

Département de géomatique appliquée
Faculté des lettres et sciences humaines
Université de Sherbrooke

TITRE DU MÉMOIRE :
**CARTOGRAPHIE PRÉDICTIVE MINÉRALE DES
DÉPÔTS DE COLTAN AU CENTRE-OUEST DE LA CÔTE
D'IVOIRE**

Par
Kassi Olivier Shaw

Mémoire présenté pour l'obtention du grade de Maître ès sciences géographiques
(M. Sc.), cheminement de type recherche en géomatique appliquée

Novembre 2019

© Kassi Olivier SHAW, 2019

Ce mémoire a été évalué par un jury composé des personnes suivantes :

Dr. Kalifa Goïta, Professeur titulaire, Université de Sherbrooke (Directeur du mémoire).

Dr. Michaël Germain, Professeur au Département de géomatique appliquée (Codirecteur du mémoire).

Dr. Yves Voirin, Ph.D. en télédétection, Chargé de cours, Département de géomatique appliquée (Examineur interne).

Stéphane Péloquin, Ph.D. en télédétection, Consultant Géologue Expert (Examineur externe).

Résumé

En matière d'exploration minière portant sur de vastes étendues géographiques, les technologies des systèmes d'information géographique (SIG), offrent une alternative très moderne et très objective pour éviter une interprétation biaisée faite par le géologue. Ces méthodes utilisent la puissance prédictive des algorithmes de classification automatique, pour modéliser et prédire la prospectivité minérale de nouvelles régions d'intérêts peu ou mal explorées. Ce travail se situe dans le cadre d'une étude de cartographie prédictive minérale (CPM) pour de nouveaux dépôts potentiels de colombo tantalite (ou coltan), au sein du Complexe géologique Hana-Lobo (Centre-ouest de la Côte d'Ivoire). Aussi, dans un contexte de difficulté d'accès à l'information géominière, un géoportail interactif est une alternative moderne pour planifier et guider des travaux de prospection complémentaires, sur les emplacements de nouvelles cibles minérales. Dans un environnement SIG, la distribution géostatistique de différents types de cibles minérales connues et la répartition spatiale de critères de preuves hydromorphologiques et structuraux sous-jacents ont été modélisées à l'aide de plusieurs algorithmes de classification automatique. Les résultats montrent que, des modèles prédicteurs basés sur des algorithmes *random forest* (RF), des machines de support vectoriel (SVM) et des K plus proches voisins (KNN) sont performants pour cartographier et prédire, à des taux satisfaisants, de nouvelles zones prospectives au sein du Complexe géologique. Cependant, les résultats de l'apprentissage automatique et la validation montrent que, les modèles prédicteurs RF et KNN sont plus précis, plus performants et robustes (à 70 % avec un coefficient Kappa de 0,6) que le modèle SVM. Le modèle prédicteur RF est utilisé pour prédire et cartographier à un taux satisfaisant, de nouvelles zones potentiellement prospectives. Les nouvelles zones prédites, occupent 58 % d'une zone de contact métamorphique, réputée favorable aux minéralisations colombo tantalifères. Un géo-portail (SIG *Web*) accessible depuis des plateformes mobiles et connectées est développé et est capable de diffuser de l'information spatiale géominière au plus près de la source.

Mots clés : Cartographie prédictive minérale - SIG *Web* – Apprentissage automatique - Exploration minière – Coltan - Côte d'Ivoire.

Abstract

In terms of mineral exploration covering large geographic areas, geographic information systems (GIS) technologies offer a very modern and objective alternative to avoid a biased interpretation made by the geologist. These methods use the predictive power of automatic classification algorithms to model and predict the mineral prospectivity of new regions of interest that are little or badly explored. This work is part of a mineral predictive mapping (MPM) study for potential new deposits of colombo tantalite (or coltan), within the Hana-Lobo Geological Complex (Central West of Côte d'Ivoire). Also, regarding exploration of locations of these new mineral targets, an interactive geoportal is a modern alternative for planning and guiding complementary prospecting work in a context of difficulty in accessing geominerary information. In a GIS environment, the geostatistical distribution of different types of known mineral targets and the spatial distribution of underlying hydro-morphological and structural evidence criteria were modeled using several automatic classification algorithms. The results show that predictive models based on random forest (RF) algorithms, vector support machines (SVM) and K nearest neighbors (KNN) are all three efficient for mapping and predicting, at satisfactory rates, new prospective areas within the Geological Complex. However, the results of machine learning and validation show that the RF and KNN predictor models are more precise, more efficient and robust than the SVM model (at 70% with a Kappa coefficient of 0.6). The RF predictor model is used to predict and map new potentially prospective areas with satisfactory rate. The new predicted zones occupy 58% of a metamorphic contact zone, known to be favorable for colombo tantaliferous mineralization. A geo-portal (Web GIS) accessible from mobile and connected platforms is developed and can disseminate geominerary spatial information as close as possible to the source.

Keywords: Mineral predictive mapping - web GIS – Machine Learning - Mining exploration - Coltan - Ivory Coast.

Remerciements

La réalisation de ce travail a été possible grâce à plusieurs personnes envers lesquelles je souhaite exprimer toute ma gratitude et mes remerciements.

J'aimerais tout d'abord remercier mes directeurs de mémoire pour leur encadrement et leurs précieux conseils, qui ont mené à la réussite de ce travail. En premier lieu, je tiens à remercier spécialement mon directeur de mémoire Prof. Kalifa Goïta, Professeur titulaire au département de géomatique appliquée, Université de Sherbrooke, qui a accepté de diriger mon travail de maîtrise. Je tiens également à remercier mon codirecteur de mémoire Prof. Michaël Germain de l'Université de Sherbrooke, qui a accepté de codiriger ce sujet.

Je tiens aussi à exprimer mes sincères remerciements aux membres du jury qui ont accepté d'évaluer mon mémoire de maîtrise.

Le financement de cette formation académique a été rendu possible grâce à la collaboration des gouvernements canadien et ivoirien, à travers le Programme de Bourse de la Francophonie (PCBF).

Finalement, je voudrais exprimer ma reconnaissance envers ma famille pour leur support moral et leurs encouragements sans faille durant mes études.

Table des matières

Résumé.....	ii
Abstract.....	iii
Remerciements.....	iv
Table des matières.....	v
Liste des figures.....	viii
Liste des tableaux.....	xi
Liste des acronymes et abréviations.....	xii
Liste des unités.....	xiii
CHAPITRE I. INTRODUCTION	1
1.1 Problématique.....	2
1.2 Objectifs et hypothèses de recherche	3
CHAPITRE II. CADRE THÉORIQUE	5
2.1 Méthode traditionnelle de prospection de placers minéralisés.....	5
2.2 Minéralisations colombo-tantalifères et pièges naturels	6
2.3 Notions de critères de preuves et de prédicteurs spatiaux dans la CPM.....	12
2.4 Techniques utilisées dans la modélisation prédictive minérale	15
2.5 Techniques de classification par apprentissage automatique.....	19
2.6 Outils d'aide à la décision pour la recherche minière	31
CHAPITRE III. CADRE EXPÉRIMENTAL	36
PRÉSENTATION DE LA ZONE D'ÉTUDE	36
3.1 Sites d'études.....	36
3.2 Données utilisées.....	46
3.2.1 Données géologiques	47
3.2.2 Données hydrologiques.....	47

3.2.3 Données morphologiques.....	49
3.2.4 Données structurales	50
3.2.5 Données de prospection minière.....	52
3.2.6 Données administratives et géominières.....	53
MÉTHODE DE TRAITEMENT DES DONNÉES	54
3.3 Méthode de discrimination des cibles minérales.....	57
3.4 Techniques d'extraction des critères spatiaux de preuves	60
3.4.1 Techniques d'extraction des paramètres hydrologiques	60
3.4.2 Techniques d'extraction des paramètres morphologiques	61
3.4.3 Techniques d'extraction des paramètres structuraux	62
3.4.4 Techniques de validation des linéaments naturels	65
3.4.5 Technique de cartographie de la densité de fracturation	66
3.5 Méthode de modélisation prédictive par apprentissage automatique.....	67
3.6 Technique de construction du jeu de données d'apprentissage.....	67
3.7 Technique d'entraînement des classifieurs automatiques	68
3.7.1 Paramètres d'apprentissage des classifieurs SVM.....	71
3.7.2 Paramètres d'apprentissage des classifieurs KNN.....	72
3.7.3 Paramètres d'apprentissage des classifieurs DT	73
3.8 Technique d'amélioration de performance d'un classifieur.....	75
3.9 Technique d'évaluation de performance d'un classifieur	76
3.10 Technique de validation et test de robustesse des modèles prédicteurs.....	79
3.11 Méthode de cartographie prédictive dans un SIG	80
CHAPITRE IV. ANALYSES ET INTERPRÉTATIONS DES RÉSULTATS	83
4.1 Discrimination des différents types de cibles minérales	83
4.2 Validation des linéaments naturels et analyse des prédicteurs spatiaux	90

4.3 Analyse des résultats d'apprentissage automatique	108
4.4 Évaluations des meilleurs modèles prédicteurs.....	115
4.4.1 Analyse des modèles en termes de précisions globales	115
4.4.2 Analyse des modèles en termes d'AUC.....	116
4.4.3 Analyse des modèles en termes de matrices de confusions	124
4.4.4 Analyse de robustesse et validation des meilleurs modèles prédicteurs.....	128
4.5 Analyse des nouvelles cartes prédictives minérales.....	131
4.6 Le géoportail interactif de l'information géominière.....	139
4.7 Cibles minérales	142
4.8 Prédicteurs spatiaux.....	143
4.9 Nouvelles zones prospectives.....	144
4.10 Modèle prédicteur final	146
CHAPITRE V. CONCLUSIONS ET RECOMMANDATIONS	149
RÉFÉRENCES BIBLIOGRAPHIQUES.....	151

Liste des figures

Figure 1 : « Tiré de Allou, (2005, p.273) » Modèle géodynamique de mise en place des pegmatites colombo-tantalifères et des roches associées d’Issia.	8
Figure 2 : « Tiré de MathWorks (1994) » Représentation schématique d’un hyperplan, de trois vecteurs de support et d’une surface de séparation de données (+,-).	20
Figure 3 : « Modifié de Sun et al. (2019, p. 34) » Organigramme d’un arbre de classification utilisant un algorithme RF.	28
Figure 4 : « Tiré de He et al. (2009, p.1267) » (a) Exemple des K-voisins les plus proches pour une observation x_i considéré ($K = 6$). (b) Création de données basée sur la distance euclidienne.	30
Figure 5: Architecture SIG Web formée avec des outils Open Source	35
Figure 6 : Situation géographique de la partie nord de la zone d'étude	36
Figure 7 : Situation géographique de la partie sud de la zone d'étude.....	37
Figure 8 : « Modifié de Yao et al. (2016, p. 857) ». Localisation du bassin versant de la Lobo	41
Figure 9 : « Modifié de Papon et al. (1971, p. 23) ». Esquisse du complexe géologique Hana-Lobo-Davo.	42
Figure 10 : Carte géologique régionale de la zone d'étude.	44
Figure 11 : « Tiré de Allou (2005, p.196) ». Architecture générale du manteau d'altération de la mine Étienne-Méguhé.	45
Figure 12: Image satellite Aster DEM (ASTGTM2_N06W007)	48
Figure 13: Mosaïque d’images multispectrales Sentinel-2B MSI L1C	50
Figure 14 : Image radar Sentinel 1A SAR IW GRD	51
Figure 15 : Imagerie haute résolution Terra Color	52

Figure 16 : Organigramme méthodologique général.	56
Figure 17 : Organigramme méthodologique d'extraction automatique de linéaments.....	63
Figure 18 : Courbe d'autocorrélation spatiale (P-value vs Z-score.)	84
Figure 19: Carte des résultats de l'analyse statistique d'autocorrélation spatiale (Site de Bémadi).....	87
Figure 20 : Carte des résultats de l'analyse statistique d'autocorrélation spatiale (Site de Brokoua)	88
Figure 21: Carte des résultats de l'analyse statistique d'autocorrélation spatiale (Site de Makua)	89
Figure 22: Carte des linéaments naturels du site de Bémadi	90
Figure 23 : Carte des linéaments naturels du site de Brokoua.....	91
Figure 24: Carte des linéaments naturels du site de Makua.....	92
Figure 25 : Rosace directionnelle régionale.....	93
Figure 26: Cartes de densité de fracturation du site de Bémadi.	94
Figure 27: Carte de densité de fracturation des sites de Brokoua et Makua.....	95
Figure 28: Cartes des profils et courbures de pentes du site de Bémadi.....	96
Figure 29 : Cartes des profils et courbures de pentes des sites de Brokoua et Makua	97
Figure 30: Cartes des degrés d'inclinaison des pentes du site de Bémadi.....	99
Figure 31: Cartes des degrés d'inclinaison des pentes des sites de Brokoua et Makua..	100
Figure 32: Cartes du relief du site de Bémadi.....	101
Figure 33: Cartes du relief des sites de Brokoua et Makua	102
Figure 34: Cartes des zones d'accumulation de flux du site de Bémadi	103

Figure 35: Cartes des zones d'accumulation de flux des sites de Brokoua et Makua	104
Figure 36: Cartes d'amplitudes et de directions des flux du site de Bémadi.....	105
Figure 37: Cartes d'amplitudes et de directions des flux des sites de Brokoua et Makua	106
Figure 38: Courbes ROC et valeurs AUC du meilleur modèle ranfom forest (RF)	117
Figure 39 : Courbes ROC et valeurs AUC du meilleur modèle KNN (pondéré-euclidienne (IDW).....	118
Figure 40: Courbes ROC et valeurs AUC du meilleur modèle SVM à noyau gaussien fin	119
Figure 41: Histogrammes comparés des performances prédictives.....	121
Figure 42: Cartes des potentialités minérales de coltan : (a) Secteur de Brokoua - (b) Secteur de Makua.....	133
Figure 43: Cartes des potentialités minérales de coltan : Secteur de Bémadi.	134
Figure 44: Carte d'occupation des zones prospectives selon les zones de contact	137
Figure 45: Vues du géoportail « Infominci ».....	140

Liste des tableaux

Tableau 1: Tableau de synthèse des données utilisées.....	46
Tableau 2 : Paramètres des algorithmes d'apprentissage automatique	69
Tableau 3 : Résultats de l'analyse statistique de Moran.....	83
Tableau 4 : Résultats des classifications automatiques avec les SVM	108
Tableau 5 : Résultats des classifications automatiques avec les KNN	110
Tableau 6 : Résultat des classifications automatiques avec les DT uniques.....	112
Tableau 7: Résultats des classifications automatiques avec les DT assemblés	113
Tableau 8: Précisions et valeurs d'AUC des trois meilleurs modèles prédicteurs.	115
Tableau 9: Composition numéraire des classes d'apprentissage.	122
Tableau 10: Matrices de confusions comparées en termes de nombre d'observation.....	125
Tableau 11: Matrices de confusions comparées en termes de TVP et TFP. Les valeurs sont en pourcentage.....	125
Tableau 12 : Matrices de confusions comparées en termes de VPP et VPN. Les valeurs sont en pourcentage.....	126
Tableau 13 : Valeurs en pourcentage des F-score comparés.	127
Tableau 14 : Composition numéraire du jeu de données de validation.	128
Tableau 15 : Matrices de confusions (en nombre et en pourcentage) des tests de validation.....	129
Tableau 16: Répartition en pourcentage des zones prospectives dans la zone de contact.	136

Liste des acronymes et abréviations

ACP	:	Analyse en composantes principales
AL	:	Aléatoire
ATHR	:	<i>Angular difference threshold</i>
AUC	:	<i>Area under curve</i>
Be	:	Béryllium
CPM	:	Cartographie prédictive minérale
DT	:	<i>Decision tree</i>
DTHR	:	<i>Distance linking threshold</i>
EW	:	<i>East-West</i>
FTHR	:	<i>Line fitting threshold</i>
GTHR	:	<i>Edge gradient threshold</i>
HH	:	<i>High High</i>
HL	:	<i>High Low</i>
KNN	:	<i>K nearest neighbors</i>
LTHR	:	<i>Curve lenght threshold</i>
LH	:	<i>Low High</i>
LL	:	<i>Low Low</i>
MNT	:	Modèle numérique de terrain
N	:	<i>North</i>
Nb	:	Niobium
NE	:	<i>North-East</i>
NS	:	<i>North-South</i>
NW	:	<i>North-West</i>
RADI	:	<i>Filter radius</i>

RF	:	<i>Random forest</i>
SVM	:	<i>Support vector machine</i>
OGC	:	<i>Open geospatial consortium</i>
ROC	:	<i>Receiver operating characteristic</i>
SGBRS	:	Système de gestion de base de données à référence spatiale
SIG	:	Système d'information géographique
SODEMI	:	Société pour le développement minier
Ta	:	Tantale
WFS	:	<i>Web feature service</i>
WMS	:	<i>Web map service</i>

Liste des unités

g/m³	:	Gramme par mètre cube
km²	:	Kilomètre carré
m	:	Mètre

CHAPITRE I. INTRODUCTION

Les ressources minières de la Côte d'Ivoire (Afrique de l'Ouest) se répartissent sur plus des deux tiers de sa superficie. Malgré l'identification de nombreuses occurrences minérales, la production minière reste essentiellement d'or et de manganèse. Au nombre des ressources potentielles, figurent près de 180 indices de minéralisation de colombo tantalite (coltan) peu ou mal explorés. Le coltan est un minéral lourd, de couleur gris-bleu, très dense, résistant à la corrosion et bon conducteur de chaleur et d'électricité. À ce titre, le tantale contenu dans ce minerai est de plus en plus utilisé dans l'industrie de l'aéronautique, pour sa résistance à la corrosion. Ce minéral est également utilisé dans la confection des équipements électroniques, pour sa bonne conductivité électrique et sa grande résistance aux températures élevées. La mise en valeur de ces gisements représente donc un enjeu stratégique pour le développement économique d'un pays. Mais pour y parvenir, il est nécessaire de collecter et d'analyser une grande quantité de données d'exploration *in situ* ou télédéetectées, en vue d'améliorer les connaissances sur les systèmes minéraux. Aussi, la mise en valeur d'un gisement minier nécessite, entre autres, de développer un outil d'aide à la décision. Cet outil doit être capable de guider des travaux de recherche minière en vue de délimiter de nouvelles zones prospectives (dans des régions d'intérêts) de façon rapide et efficiente. Cependant, très peu de travaux de recherche scientifique ont porté sur l'étude des systèmes minéraux du coltan en Côte d'Ivoire. L'intérêt que suscite l'utilisation des SIG en matière de recherche minière, dans un contexte de données d'exploration incomplètes, incertaines ou imprécises, justifie ce projet de recherche en géomatique appliquée. Ce projet s'attaque au défi immense qu'est celui d'identifier de nouveaux gisements à partir d'observations et de données incomplètes, par l'application de nouvelles techniques de cartographie prédictive minérale (CPM). Ce projet vise également à mettre en place une nouvelle infrastructure technologique innovante et conviviale, utilisable pour le partage de l'information géominière et la planification de travaux d'exploration sur de nouvelles cibles minérales potentielles.

1.1 Problématique

Délimiter de nouvelles zones favorables à l'exploration minérale sur de vastes étendues géographiques nécessite la collecte et l'analyse d'une grande quantité de données *in situ* ou télédéetectées très onéreuses à acquérir. Des méthodes traditionnelles de prospection de placers sont appliquées par le géologue en vue d'identifier de nouvelles zones prospectives potentiellement minéralisées et à valeurs économiques. La démarche consiste à compiler l'ensemble de l'information géominière disponible, à étudier le système de minéralisation en présence, à traiter, interpréter et modéliser l'information (de la façon la plus objective possible) en vue de délimiter de nouveaux prospects accessibles. Mais le manque de connaissances et de données d'exploration acquises sur les systèmes minéraux du coltan, notamment au centre-ouest de la Côte d'Ivoire, engendre des difficultés pour analyser, modéliser et prédire la distribution géostatistique de nouvelles cibles minérales. Dans un tel contexte, les technologies des systèmes d'information géographique (SIG), offrent une alternative très moderne et très objective pour éviter une interprétation biaisée faite par le géologue. Ces méthodes utilisent la puissance prédictive des algorithmes de classification automatique, pour modéliser et prédire la prospectivité minérale de nouvelles régions d'intérêts peu ou mal explorées. Ces nouvelles techniques sont utilisées pour modéliser des prédicteurs spatiaux, sans avoir recours aux jugements d'experts, tout en se servant d'emplacements d'occurrences connues, comme points de contrôle. Cependant, la procédure la plus critique dans la modélisation d'une prospectivité minérale est la sélection de critères de ciblage appropriés et l'application de techniques innovantes et robustes pour dériver les caractéristiques probantes de ces critères (Joly *et al.*, 2012). En matière de recherche minière, ces critères sont nécessaires pour étudier les relations spatiales existantes entre des occurrences minérales connues et des indicateurs probants de minéralisation. Les cartes de potentialités minérales résultantes peuvent alors participer aux prises de décisions des gestionnaires, en matière de développement durable, d'aménagement du territoire et de gestion optimale de ressources naturelles. D'autre part, l'estimation de nouvelles ressources minérales (ou réserves) au sein d'une zone cible potentiellement minéralisée nécessite de planifier de nouveaux travaux de prospection plus approfondis. Pour y parvenir, le géologue a de plus en plus recours à un environnement SIG pour l'aider à collecter, traiter, analyser et interpréter de grandes quantités de données

géospatiales, hétéroclites, disparates et parfois obsolètes. C'est pourquoi, de nos jours, des cartes interactives connectées sont de plus en plus utilisées par le géologue en vue d'accéder à un éventail de données géominières numériques, au plus près de la source. Dans un contexte de pays en voie de développement (comme la Côte d'Ivoire) l'accès en ligne à un ensemble d'information géominière régionale ou locale (disponible ou à acquérir) et au plus près de la source, est quasi inexistante. Dans ce cas, la mise en place d'une application cartographique interactive, capable de compiler et de diffuser une information géominière à partir d'une plateforme mobile et connectée, peut s'avérer être un outil d'aide innovant pour optimiser la planification de nouvelles campagnes de terrain, au sein de nouvelles zones cibles prospectives.

1.2 Objectifs et hypothèses de recherche

L'objectif principal de ce travail est de cartographier la prospectivité de nouvelles zones potentiellement favorables aux dépôts minéraux de coltan, au sein du Complexe géologique Hana Lobo, en utilisant une approche de combinaison d'indicateurs spatiaux dans un environnement SIG et de fournir un géoportail capable de diffuser des données géospatiales complémentaires et utilisables pour planifier des travaux (plus approfondis) de prospection minière.

Pour y parvenir, trois objectifs spécifiques se dégagent :

- 1) analyser dans un environnement SIG, la distribution géostatistique des données utilisées pour la prospection minérale du coltan, en considérant les données hydromorphologiques et structurales comme critères spatiaux de preuves sous-jacents ;
- 2) modéliser la distribution géostatistique des cibles minérales en fonction des critères de preuves, afin de prédire de nouvelles zones prospectives et ;
- 3) réaliser une application cartographique SIG *Web* se servant de l'approche développée, pour permettre guider la conduite de travaux de prospection plus approfondis.

Pour atteindre ces objectifs spécifiques, le projet s'appuie sur les hypothèses suivantes :

- 1) les données géologiques disponibles (utilisées) permettent d'identifier l'emplacement de nouvelles cibles ;
- 2) l'analyse statistique d'agrégation spatiale de Moran permet d'expliquer le mode de distribution des données de prospection minérale et de discriminer différents types de cibles minérales ;
- 3) les algorithmes de classification par apprentissage automatique permettent de modéliser la distribution géostatistique des cibles minérales, en fonction de critères spatiaux de preuves et de prédire la prospectivité minérale de nouvelles régions d'intérêts;
- 4) une application cartographique SIG *Web* est un outil interactif innovant pour démocratiser l'accès à une information géospatiale minière au plus près de la source et utilisable pour optimiser la planification de nouveaux travaux de prospection de zones cibles potentiellement prospectives.

CHAPITRE II. CADRE THÉORIQUE

Cette partie du travail passe en revue les méthodes traditionnelles utilisées pour la prospection de placers minéralisés, présente de façon contextuelle le système de minéralisation étudié et les nouvelles notions et techniques utilisées dans la littérature, en matière de cartographie prédictive minérale et d'apprentissage automatique. Elle se termine par la présentation des architectures SIG *Web* utilisables pour diffuser de l'information de géologie minière.

2.1 Méthode traditionnelle de prospection de placers minéralisés

Le principe des méthodes de prospection de placers est de sélectionner parmi différents types de placers connus, ceux qui possèdent une valeur économique, c'est-à-dire ceux dont l'exploitation donnera un bénéfice suffisant (Anthoine, 1941). Au fil des années, se sont développées et perfectionnées plusieurs méthodes de prospection de gîtes secondaires. Ces méthodes sont applicables sur de vastes zones favorables aux dépôts de graviers (potentiellement minéralisés) alluvionnaires, colluvionnaires ou éluvionnaires. Il s'agissait de perfectionner (suivant le contexte géomorphologique et le type de minéralisation en présence) différentes techniques empiriques développées par d'éminents géologues-prospecteurs, en vue d'extraire un maximum de métaux piégés au niveau de bancs de graviers.

De façon générale, toute technique d'exploration de placers commence par une prospection volante. Il s'agit (généralement à vue) de positionner, aux emplacements susceptibles d'être favorables au piégeage de métaux, un ensemble de puits de reconnaissance. Puis, d'analyser (au laboratoire ou sur le terrain) les teneurs minéralogiques ou géochimiques en métaux contenues dans les couches graveleuses échantillonnées. Les résultats d'analyses obtenus sont alors positionnés sur une carte de prospection, intégrant des levés de surfaces géologiques, topographiques ou hydrographiques, en vue d'identifier de nouvelles zones d'intérêts potentiellement prospectives. L'exploration de ces nouvelles zones se fait généralement au moyen d'une grille de prospection à maille régulière positionnée autour des puits les plus minéralisés.

Ainsi, le géologue peut évaluer (à la suite de nouvelles analyses géologiques, minéralogiques ou géochimiques) l'étendue de la minéralisation tant en surface qu'en profondeur.

Avec le développement des nouvelles techniques de télédétection et de SIG, des résultats de traitement de données géoscientifiques spatiales, acquises par levés géophysiques, par traitement d'image radar ou optique, et récemment par lidar, sont superposés à des données de prospection (volante ou maillée) pour réaliser des cartes de potentialités minérales régionales ou locales. Ces cartes ou ébauches de cartes sont utilisables pour planifier et guider des travaux de prospections complémentaires sur des sites peu ou mal explorés.

2.2 Minéralisations colombo-tantalifères et pièges naturels

Dans un contexte de données d'exploration minière incomplètes et de dimensionnalité spatiale élevée, une bonne connaissance du système de minéralisation étudié est une étape importante. Cela passe nécessairement par une bonne compréhension du contexte métallogénique et/ou paragénétique du type de dépôts en présence. Dans le cadre d'une CPM, cette étape doit être réalisée dans le but de cartographier et prédire de façon la plus efficiente possible, de nouvelles zones prospectives dans des régions peu ou mal explorées. Par conséquent, il s'avère primordial d'identifier le rôle et d'évaluer la pertinence de différents facteurs inductifs ou explicatifs, susceptibles d'influencer (de façon directe ou indirecte) les processus de minéralisation et/ou de formation des dépôts. Ces facteurs peuvent être alors traduits suivant une approche analytique dans un environnement SIG et être utilisés comme critères de preuves. Les paragraphes suivants dressent un état des connaissances acquises sur des systèmes de minéralisation colombo-tantalifère en Afrique, en Côte d'Ivoire et particulièrement au sein du Complexe géologique Hana-Lobo.

Minéralisations colombo-tantalifères d'Afrique

Plusieurs systèmes de minéralisations colombo-tantalifères du continent africain ont fait l'objet d'étude cette dernière décennie (Melcher *et al.*, 2015). Notamment, les minéralisations à éléments rares des formations granitiques et dans les pegmatites, à l'instar des minéralisations du bouclier de Man en Sierra Leone, du craton du Congo en République démocratique du Congo, de la ceinture de Kamativi au Zimbabwe, du plateau de Jos au

Nigéria, de la province du Alto Ligonha au Mozambique, du bouclier arabo-nubien et de la vallée de la tantalite au sud de la Namibie. Mais très peu d'études ont porté sur les minéralisations de type Niobium-Tantale-Lithium (Nb-Ta-Li) d'Afrique de l'Ouest et particulièrement en Côte d'Ivoire.

En Afrique, il ressort que la mise en place des minéralisations connues dans les pegmatites et les granites à éléments rares (Ta-Nb), est liée soit aux épisodes syn orogéniques tardifs (craton ouest-africain, craton zimbabwéen), soit aux épisodes post orogéniques (ceinture de Kibara, ceinture de Damara, granites anciens du Nigeria, ceinture d'Adola en Éthiopie) ou soit aux épisodes tectoniques et magmatiques anorogéniques (granites récents du Nigéria) (Melcher *et al.*, 2015).

Minéralisations colombo-tantalifères de Côte d'Ivoire

Les problématiques de la présence régionale des minéralisations en Nb-Ta au sein des granites et des pegmatites et plus localement, celle de la distribution spatiale des teneurs de coltan au sein des gisements connus, ont été abordées par quelques travaux de prospection minière.

Les premières études géologiques et minéralogiques ont tenté d'expliquer l'origine et la présence de cette minéralisation (Rombach, 1960). Les secondes ont tenté de comprendre le mode de distribution spatiale et de concentration des teneurs de coltan dans les alluvions, les colluvions et les éluvions, occupant les zones de flats minéralisés (Cruys, 1965; Adam, 1968; (Papon *et al.*, 1971). D'autres se sont focalisées sur l'origine métallogénique des minéralisations primaires du colombo-tantalite dans la région nord du Complexe géologique Hana-Lobo (Feybesse *et al.*, 1990; Vidal *et al.*, 1992 ; Allou, 2005).

Il ressort de cela que des concentrations métallifères d'importances variables, notamment le colombo-tantalite, le béryl, le lithium et l'or, ont accompagné la mise en place des formations granitiques et des pegmatites, survenues lors de différents épisodes orogéniques. En Côte d'Ivoire, les minéralisations à éléments rares (Ta-Nb) les plus connues sont liées à l'orogénèse éburnéenne et appartiennent à la dorsale de Man (craton ouest-africain). La figure 1 présente de façon schématique l'histoire géologique régionale de mise en place du système de minéralisation étudié.

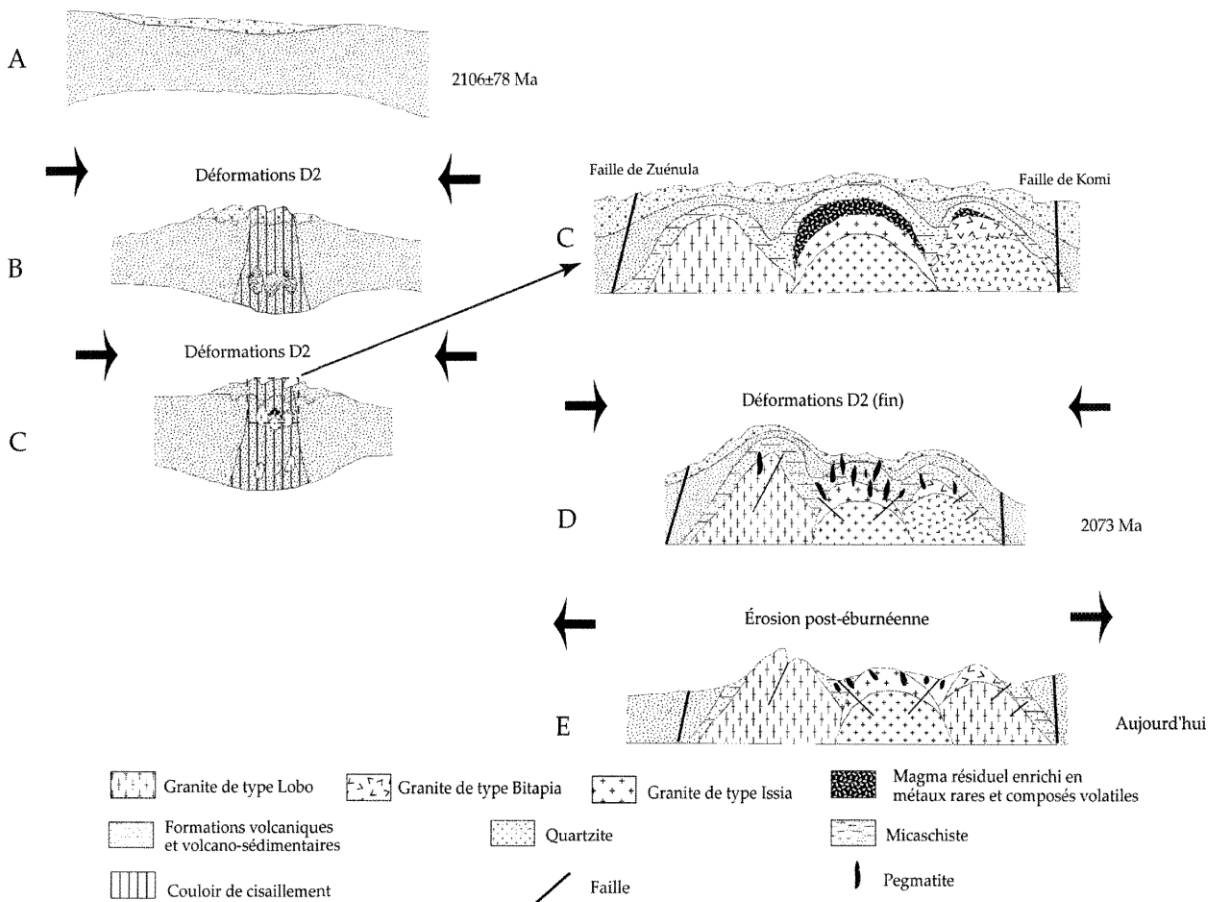


Figure 1 : « Tiré de Allou, (2005, p.273) » Modèle géodynamique de mise en place des pegmatites colombo-tantalifères et des roches associées d'Issia.

La première phase (A) est marquée par un régime de tectonique transcurrente D2. Ce régime tectonique est à l'origine des grandes failles régionales observables sur la zone d'étude (faille de Zuénoula et de Kami). Cette déformation transcurrente D2 est accompagnée par plusieurs fractures conjuguées (les fractures de Riedel et les fentes de tractions). Cette déformation a affecté à partir de 2106 ± 78 Ma, les formations sédimentaires et volcan sédimentaires, issues de l'érosion des topographies anciennes, générées par les déformations antérieures D1 (Feybesse *et al.*, 1990 ; Vidal *et al.*, 1992). La phase (B) est marquée par l'apparition de trois poches de fusion de la croûte continentale en profondeur, suivant une direction de cisaillement principale NS à NNW-SSE. La phase (C) se caractérise par la granitisation progressive vers la surface, par l'intermédiaire de couloirs de décrochement. Les granites venant de masses chaudes génèrent sur leurs

bordures d'importantes auréoles de métamorphismes de contact dans les micaschistes. La phase (D) se caractérise par la formation de magmas résiduels (riches en composés volatils et en métaux rares (Nb-Ta-Li-Au-Sn...)), qui se sont mis en place à la suite d'une forte cristallisation fractionnée (Allou, 2005).

Minéralisations colombo-tantalifères du Complexe géologique Hana-Lobo

Au sein de ce Complexe, la présence du colombo-tantalite est généralement associée aux pegmatites. Dans ces roches, essentiellement constituées de quartz, de feldspaths (fréquemment de la microcline perthitique) et de micas (muscovite surtout), s'accumulent sous certaines conditions, des éléments rares comme le niobium (Nb), le tantale (Ta), le lithium (Li), l'uranium (U), le thorium (Th), les lanthanides, etc.). L'accumulation de ces éléments rares dans la roche peut former des minéraux dits pneumatolytiques tels que le colombo-tantalite, le béryl, la tourmaline et donner lieu à des gîtes miniers exploitables.

À l'échelle du Complexe géologique et concernant la problématique sur l'origine et la présence des dépôts connus, Rombach (1960) affirmait qu'il se peut que les colombo-tantalites détritiques proviennent en partie de filons quartzeux pneumatolytiques recoupant les corps pegmatitiques. Pour Cruys (1965) et plus tard Adam (1968), les granites ont été le siège d'importants processus pegmatitiques, pneumatolytiques et hydrothermaux qui ont conduit à la mise en place de filons et de roches d'altération hydrothermales de natures diverses. Il s'agit des pegmatites, des aplites, des tourmalinites, des quartz et des greisens. Adam, (1968) renchérit en affirmant que la présence des colombo-tantalites dans cette zone est une caractéristique régionale, beaucoup plus qu'une caractéristique métallogénique des pegmatites. Toutefois, selon Allou, 2005, la localisation des placers à la périphérie des massifs de pegmatites granitiques ne fait pas douter de son origine pegmatitique. Car, à l'échelle des placers, les gisements sont des concentrations détritiques en minéraux de Tantale (Ta) et de Niobium (Nb) issus de la désagrégation de roches primaires situées au sommet et sur les flancs des massifs, situés eux-mêmes, en amont des zones alluvionnaires des placers minéralisés. En effet, il est avéré de nos jours que la plupart des dépôts connus ont été mis en évidence dans des placers.

Les gisements colombo-tantalifères d'Issia (centre-ouest de la Côte d'Ivoire) localisés dans la partie nord du Complexe géologique Hana-Lobo, ont fait l'objet d'une exploitation par la Société pour le développement minier de la Côte d'Ivoire (SODEMI).

Concernant la partie centre du Complexe géologique (Zone de Grand-Zatry et Oureyo), les travaux réalisés par Papon *et al.* (1971) ont mis en évidence plusieurs indices minéraux de coltan. Ces indices, répartis sur l'axe routier Oureyo-Dinyago, sont des concentrés alluvionnaires de colombo-tantalites identifiés dans des graviers, avec des teneurs, variant entre 15 et 60 g/m³. Aussi, quelques affleurements de pegmatites ont été mis en évidence dans cette région, à l'occasion de quelques travaux de terrassement routier. Cependant, aucun gisement connu de coltan n'a à ce jour été identifié dans cette partie du complexe, réputée très aurifère. Des travaux très récents de prospection régionale, entrepris dans cette zone, dans le cadre de notre campagne de terrain en 2018, confirment toutefois la présence de graviers minéralisés et de pegmatites, au niveau des sous-bassins versants prospectés.

Placers et pièges naturels

Les placers sont des zones de concentration de minéraux résultant de processus géodynamiques externes, comme l'altération, l'érosion, le transport et le dépôt de sédiments. De même que l'or, le diamant ou la cassitérite, le colombo-tantalite peut s'accumuler dans certaines conditions et donner lieu à la formation de gîtes exploitables. En effet, dans un contexte régional tropical humide, les conditions météoriques très poussées sont le siège de phénomènes d'altération importants, qui favorisent la désagrégation de la roche primaire minéralisée. Dans ces conditions, les minéraux très résistants à l'altération météorique, comme le quartz, les minéraux niobo-tantalifères et le béryl, sont susceptibles de rester intacts en place et de s'accumuler. Ce qui contribue à un enrichissement des graviers dans des zones topographiques favorables à leur accumulation.

En effet, au cours du transport mécanique des sédiments, les minéraux lourds sont susceptibles de se déposer au contact du socle, dans des endroits présentant de fortes irrégularités topographiques. Et, pour que le dépôt de ces minéraux lourds soit constant au même endroit et donne des gîtes exploitables, il faut que les irrégularités topographiques soient fixes dans l'espace et dans le temps (Bouladon, 1964).

Ces endroits constituent par conséquent, des zones de pièges naturels favorables à l'accumulation de graviers minéralisés. De telles zones de pièges naturels ont été déjà mises en évidence dans la partie nord du complexe géologique par les travaux de Cruys, (1965). Des teneurs exploitables de 153 à 2085 g/m³ y ont été ainsi découvertes. Ces teneurs ont été obtenues au moyen de puits positionnés en travers des flats, dans des placers alluvionnaires, éluvionnaires et colluvionnaires de la région d'Issia.

À l'échelle du gisement et concernant la problématique de la répartition spatiale des teneurs en coltan, notamment dans les placers alluviaux, colluviaux et éluviaux ; les travaux de Cruys, (1965) ont montré que les réserves identifiées se répartissaient très irrégulièrement dans les vallées et que la concentration des teneurs était très irrégulièrement distribuée dans les graviers, même sur de très courtes distances. De plus, les profils topographiques montraient que les fortes accumulations du minerai n'étaient pas toujours en rapport strict avec les creux des microreliefs des *bedrocks*, malgré une tendance à la concentration dans les couches de graviers superposées aux cuvettes du socle.

En outre, dans la région nord du Complexe, des travaux plus récents (2015) de prospection minière, réalisés par la SODEMI, ont mis en évidence des teneurs en concentré de coltan, allant de 50 à 1157 g/m³, dans différents placers alluviaux, colluviaux et éluviaux (secteur de Bémadi, Brokoua et Makua). Ces minéralisations ont été découvertes au niveau des vallées, des flancs de collines et au sommet des massifs (pénéplaines) dominant les bassins versants prospectés.

Sur la base de cette revue littéraire et des travaux de prospections récents menés au sein du Complexe géologique Hana-Lobo, il faut retenir que le système de minéralisation primaire étudié s'est mis en place dans les granites et les pegmatites au cours des différents épisodes orogéniques qui ont marqué la région. Les dépôts secondaires, faisant l'objet de notre étude, sont de type placer et de nature détritique.

Au plan morphologique, plusieurs dépôts connus se localisent dans les zones alluvionnaires, colluvionnaires et éluvionnaires, situées dans les vallées, les flancs et au sommet de massifs dominants les bassins versants.

Au plan hydromorphologique, il faut retenir que les placers hébergeant les minéralisations de Nb-Ta de la région, se sont formés à la suite de plusieurs processus

hydromorphologiques, marqués par d'importantes phases de drainage et d'alluvionnement, subséquentes à une phase de fragmentation et d'usure des roches filoniennes (altération hydrothermale) de natures diverses, comme les pegmatites, les aplites, les tourmalinites, les quartz et les greisens (Cruys, 1965; Adam, 1968).

Sur la base de ces observations de terrain et des connaissances acquises dans la littérature, cette étude se propose de contribuer à résoudre la problématique posée sur la distribution spatiale des teneurs en coltan à l'échelle des gisements de la région d'Issia. Notre étude entend également identifier de nouvelles zones prospectives potentielles, dans la région Centre, peu ou mal explorée du Complexe géologique (Secteurs de Grand-Zatry et Oureyo). Mais pour y parvenir, il est nécessaire de traduire ces différentes connaissances minéralogiques, géologiques, topographiques, morphologiques et hydrographiques, en paramètres quantifiables et mesurables (comme critères de preuves) pour réaliser une étude de CPM dans un environnement SIG. Dans cette optique, plusieurs approches analytiques ont été identifiées à travers nos lectures.

2.3 Notions de critères de preuves et de prédicteurs spatiaux dans la CPM

Notion de critères de preuves

Pour traduire des connaissances minéralogiques, géologiques, topographiques et hydromorphologiques en paramètres mesurables et quantifiables, nous sommes partis du constat que, la lithologie, la topographie et les facteurs hydromorphologiques peuvent influencer la quantité, la qualité et la distribution spatiale de sédiments minéralisés à l'échelle d'un bassin versant (Sable *et al.*, 2006). En effet, Zuluaga *et al.* (2017) affirment que les sédiments fluviaux peuvent transporter des minéraux mobilisés, à partir de pentes adjacentes. Ces minéraux accumulés en aval sont représentatifs de la composition lithologique et minéralogique des sols en amont des bassins versants. En effet, ces sédiments sont le produit composite de l'érosion et des intempéries. Leur nature et leur composition minéralogiques reflètent la géologie du substrat rocheux (affleurant ou sous-jacent), la couverture des morts-terrains et les minéralisations métallifères qui ont cours au sein du bassin versant (Sadeghi *et al.*, 2015).

L'une des techniques utilisées dans la littérature pour cartographier la distribution spatiale de sédiments de ruisseaux minéralisés est l'analyse géochimique ou minéralogique d'échantillons prélevés à l'échelle du bassin versant. La littérature propose à cet effet, une approche analytique dans un environnement SIG, pour délimiter des zones potentiellement favorables à l'accumulation de ces sédiments minéralisés. Cette procédure consiste à analyser et à interpréter la distribution spatiale des données de géologie minière indicatrices de minéralisation. L'interprétation peut se réaliser dans un SIG en fonction de différents critères hydromorphologiques et structuraux explicatifs des processus de formation, d'altération hydrothermale, de transport et de dépôt de sédiments sur une surface géographique donnée. Les facteurs hydromorphologiques sont généralement mis en évidence, par une analyse hydrologique d'un modèle d'élévation numérique (MNE) de la zone d'étude. Les cartes résultantes peuvent alors être utilisées pour identifier des zones favorables à l'accumulation de dépôts sédimentaires (Sadeghi *et al.*, 2015; Zuluaga *et al.*, 2017).

Cette approche peut être utilisée dans le cadre de notre étude pour identifier de nouvelles zones favorables à la formation de pièges naturels, qui sont eux-mêmes potentiellement favorables à la concentration de minéraux lourds au niveau des placers. Il s'agira donc d'utiliser un SIG, pour cartographier la distribution spatiale de différents critères spatiaux de preuves, en vue de déceler des zones d'anomalies caractéristiques et spécifiques au phénomène étudié.

L'ensemble des connaissances acquises sur le système minéral tend à mettre en exergue une certaine influence de facteurs hydromorphologiques et structuraux sur les processus de formation et de géolocalisation des dépôts détritiques connus au sein des bassins versants du Complexe géologique Hana-Lobo. Ces facteurs sont susceptibles de contrôler les processus d'érosion et de transport de sédiments et peuvent par conséquent, être utilisés comme critères de preuves de favorabilité ou comme prédictors spatiaux dans une étude de CPM.

En effet, dans le cadre des opérations de recherche minière et de délimitation de placers riches en minéraux lourds, comme le diamant, le colombo-tantalite ou le béryl, plusieurs types de prédictors spatiaux de minéralisation sont utilisés dans la littérature, pour

localiser des zones favorables à leur mise en place. Dans la plupart des cas, ces prédicteurs sont basés sur des critères de preuves liés à l'environnement géologique, géomorphologique ou à l'aspect des minéraux accompagnateurs (Jacob *et al.*, 1999).

En revanche, dans le cadre de la prospection de minéraux fortement malléables et aplatissables comme l'or, les indicateurs sont plutôt basés sur la nature morphologique et chimique de ces minéraux (Allou, 2005). Dans ces cas, les critères spatiaux de preuves utilisés sont généralement basés sur des données lithologiques, géologiques, géochimiques ou géophysiques.

Cependant, il n'y a pas de façon spécifique d'évaluer la signification minéralogique d'un indicateur spatial en fonction de sa nature thématique. Car chaque cas de modélisation prédictive minérale est unique, en fonction du type de gisement minéral prospecté, de la situation géographique et du contexte climatique de la zone d'étude. Chaque étude tient compte de la qualité et de la quantité de données collectées au sol et/ou détectées à distance (Carranza *et al.*, 2008). Par conséquent, le critère le plus objectif, pour sélectionner des indicateurs de minéralisation pertinents à utiliser comme prédicteurs spatiaux, serait une bonne interprétation des connaissances acquises sur le contexte géologique local, sur les modèles génétiques et les facteurs susceptibles d'influencer et/ou d'expliquer les différents processus de concentration de la substance ou du minerai recherché, dans un espace géographique donné.

Dans la littérature, plusieurs types de données thématiques ont été utilisées pour réaliser des études de CPM. Ces études ont fourni dans la majorité des cas des résultats satisfaisants. Par exemple, des données géologiques (formations minéralisatrices, formations encaissantes et failles) ont été sélectionnées comme indicateurs spatiaux, sur la base des liens métallogéniques existants entre des sources métalliques, des voies de transport de fluides magmatiques et des zones d'accumulation de minéraux. Cette étude a permis de réaliser une modélisation SIG à deux dimensions (2D) des gisements orogéniques et intrusifs dans les provinces géologiques Arunta, Musgrave et Gascoyne de l'ouest de l'Australie (Joly *et al.*, 2015). Des études de modélisation prédictive de dépôts de fer de type *skarn* du District sud-ouest du Fujian en Chine (Xiong *et al.*, 2018), ont utilisé des formations de granitoïdes et de roches carbonatées, des emplacements de failles, des

variables géochimiques caractéristiques des altérites et des anomalies aéromagnétiques comme prédicteurs spatiaux. Cela a permis de prédire, cartographier et délimiter de nouvelles zones prospectives autour de 95 % des dépôts connus. Par ailleurs, un ensemble de données d'exploration, comprenant des données géologiques, des données géochimiques de sédiments fluviatiles, des données de géophysiques aéroportées et des données de télédétection multi spectrales, a été utilisé comme prédicteurs spatiaux dans un SIG, pour modéliser et prédire à 77 % des occurrences de gisements porphyriques de cuivre-or et d'oxyde de cuivre-or dans les régions de Dalli et Aftabru au centre de l'Iran (Asadi *et al.*, 2015).

Sur la base de cette revue littéraire, et tenant compte du rôle et de l'influence probable de certains facteurs sur la mise en place des placers minéralisés en coltan, plusieurs indicateurs de minéralisation peuvent être alors définis et utilisés dans le cadre de cette étude. Ces facteurs peuvent être cartographiés dans un SIG et être utilisés comme prédicteurs spatiaux de zones favorables à la formation des dépôts étudiés.

Dans le cas du système de minéralisation étudié, il s'agit, au niveau hydromorphologique, des facteurs susceptibles de contrôler les processus de drainage et de dépôts de minéraux lourds au niveau des bassins versants. Au niveau topographique, il s'agit des zones de vallées, de flancs et de sommet de massifs dominant les bassins versants où plusieurs dépôts connus ont été observés, prospectés et/ou exploités. Au niveau morphostructural, il apparaît que dans un contexte de dépôt sédimentaire, certaines zones d'irrégularités topographiques sont susceptibles de constituer des zones favorables au piégeage de minéraux lourds sous certaines conditions. Au niveau structural et tectonique, le modèle génétique régional considéré, montre que le fluide minéralisateur dérive de la cristallisation du granite et sa migration se fait par pulsations successives le long de fractures de Riedel (R') et de fentes de traction. L'implication des systèmes de diaclases, de failles et de fractures dans le processus de mise en place des minéralisations primaires suggère la prise en compte de ce facteur dans le cadre de cette étude.

2.4 Techniques utilisées dans la modélisation prédictive minérale

La littérature propose plusieurs techniques utilisables pour parvenir à combiner plusieurs critères spatiaux de preuves dans le cadre d'une étude de CPM. Il s'agit généralement de

produire un modèle mathématique spatial capable de prédire, en termes de probabilité, la présence ou l'absence du phénomène étudié sur une étendue géographique spécifique.

Plusieurs approches sont utilisées pour modéliser des systèmes minéraux et évaluer des probabilités d'occurrences minérales au sein d'une région d'intérêt (Porwal *et al.*, 2010; Joly *et al.*, 2015). Les méthodes proposées dans la littérature sont généralement classées en deux catégories.

La première regroupe les méthodes axées sur la connaissance, qui utilisent le jugement d'experts pour attribuer des poids probants aux couches de preuves (Bonham-Carter, 1994; Carranza *et al.*, 2008), afin de prédire la prospectivité minérale de nouvelles régions d'intérêts peu ou mal explorées. Dans ce cas, la pondération des caractéristiques probantes nécessite une bonne compréhension des processus minéraux et des relations existantes, entre des facteurs probants et le type de gisement minéral recherché. Ce qui nécessite, généralement dans ces cas, l'acquisition d'une grande quantité d'information de géologie minière *in situ* ou télédétectées.

La seconde regroupe les méthodes axées sur les données. Elles utilisent les emplacements d'occurrences minérales connues comme site d'entraînement, pour modéliser des critères d'identification de prospectivité (Carranza *et al.*, 2016). Ces méthodes se basent sur une estimation d'associations spatiales entre différentes mesures quantitatives réalisées au sein des occurrences, pour attribuer un poids à chaque couche de preuve informative (Carranza, 2011). Ces méthodes sont utilisées également dans le but de prédire la prospectivité minérale de nouvelles régions d'intérêts peu ou mal explorées.

À ces deux catégories peuvent s'ajouter une approche hybride qui utilise à la fois les emplacements d'occurrences minérales connues et les opinions d'experts, dans l'attribution de poids probants aux indicateurs de minéralisations (Porwal *et al.*, 2004; Yousefi *et al.*, 2015).

De façon pratique, l'application de ces méthodes de modélisation spatiale tombe généralement dans un spectre entre le ciblage conceptuel et le ciblage empirique. En effet, un ciblage conceptuel tentera de comprendre le processus contrôlant la distribution spatiale

de la substance d'intérêt. Notamment, le processus de combinaison pour créer un dépôt économiquement rentable, en vue de prédire une zone d'accumulation favorable.

En revanche, le ciblage empirique se focalisera plutôt, sur la reconnaissance des modèles spatiaux à partir des contrôles géologiques connus et utilisables comme guide de prospection.

Cependant, dans certains cas, ces approches sont souvent complémentaires et peuvent être employées en tandem (McCuaig *et al.*, 2017). Notamment, lorsque les relations spatiales existantes entre des cibles minérales et des indicateurs de minéralisations sont complexes et non linéaires, ou dans un contexte de manque de connaissances acquises sur la distribution géostatistique d'indicateurs de minéralisation. Dans de tels cas, les données de prospection minérale peuvent être utilisées pour modéliser la distribution spatiale d'un type de gisement particulier, en vue d'identifier de nouvelles zones prospectives au sein d'un même complexe géologique (Carranza, 2011).

Pour pallier le manque de données acquises dans le cadre de la prospection minérale de vastes étendues géographiques, les études ont de plus en plus recours à des algorithmes d'apprentissage automatique pour modéliser des systèmes minéraux et prédire la prospectivité de nouvelles régions d'intérêts. En effet, cette dernière décennie, avec le développement des algorithmes d'apprentissage automatique, des modèles basés sur des fonctions de régressions logistiques (RL) (King *et al.*, 2001; Xiong *et al.*, 2018), de réseaux bayésiens (Abedi *et al.*, 2012a), de machines à supports vectoriels (SVM) (Yu *et al.*, 2012; Shabankareh *et al.* 2017), des systèmes de réseaux neuronaux (RNA) (Noack *et al.*, 2014; Zhao *et al.*, 2016) ou des arbres de décisions (DT) et notamment la forêt aléatoire (RF) (Hariharan *et al.*, 2017; Kereszturi *et al.*, 2018; Miraki *et al.*, 2019) sont de plus en plus utilisés en matière de CPM. Ces modèles prédictifs ont permis de cartographier et de prédire la prospectivité de nouvelles régions d'intérêts à des taux satisfaisants. Notamment, avec l'utilisation de la puissance d'apprentissage des RNA. Plusieurs travaux utilisant des produits de systèmes experts comme les machines Boltzmann restreintes (Chen *et al.*, 2014; Chen, 2015) ou des systèmes d'inférences neuro-floues (Porwal *et al.*, 2004; Zuo *et al.*, 2011; Abedi *et al.*, 2012a) ont montré le fort potentiel des réseaux de neurones artificiels à modéliser de façon prédictive la prospectivité de nouvelles cibles minérales.

Mais la limite d'application de cette technique pour notre projet est que ces modèles exigent un grand nombre de données d'entraînement pour l'apprentissage des algorithmes (Abedi *et al.*, 2017). Autrement, leurs performances s'en trouvent dégradées (Asadi *et al.*, 2015). De plus, les processus d'apprentissage requièrent l'estimation de valeurs de nombreux paramètres pouvant avoir un impact important sur la robustesse du modèle final (Rodriguez-Galiano *et al.*, 2015).

C'est pourquoi de nouvelles approches non paramétriques de classification automatique par noyau sont de plus en plus utilisées pour résoudre des problèmes complexes dans divers domaines de géosciences. Ces nouvelles approches ont été utilisées dans plusieurs études avec un résultat satisfaisant. Elles ont été utilisées notamment, dans une étude de classification de la couverture des sols (Otakei *et al.*, 2010), pour la cartographie et l'évaluation de la sensibilité des sols aux inondations (Tehrany *et al.*, 2015), pour une étude des glissements de terrain (Peng *et al.*, 2014) ou encore dans la prédiction d'unités hydrauliques (Ali *et al.*, 2013) ou de la biomasse de la forêt (Nguyen *et al.*, 2018), mais également pour la classification d'unité lithologique (Yu *et al.*, 2012).

En matière de CPM, ces nouvelles approches automatiques non paramétriques ont maintes fois, permis de décrire, d'expliquer ou de prédire à partir d'une série de variables, l'appartenance d'un ensemble d'observations à des groupes prédéfinis (ou classes) . Généralement, la technique consiste à construire une fonction de classement (ou une règle d'affectation) qui permet de prédire le groupe d'appartenance d'un individu à partir des valeurs prises par des variables prédictives. Dans ce sens, plusieurs travaux utilisant des algorithmes basés sur les SVM (Zuo *et al.*, 2011; Yu *et al.*, 2012; Granek *et al.*, 2015; Shabankareh *et al.*, 2017) , les K plus proches voisins (KNN) (Nguyen *et al.*, 2018) et les RF (Carranza *et al.*, 2015, 2016; Rodriguez-Galiano *et al.*, 2015; Hariharan *et al.*, 2017) ont montré de bonnes performances pour la prédictivité minérale de nouvelles régions peu ou mal explorées.

En outre, ces nouvelles approches présentent plusieurs avantages. En l'occurrence, le fait de nécessiter moins de données d'apprentissage augmente en conséquence le degré d'automatisation des classificateurs, tout en les rendant plus facilement paramétrables (Bater *et al.*, 2009). Ces « classifieurs » non paramétriques sont également utilisés en raison

de leur simplicité ou de leur « interprétabilité » ou, en raison du faible coût en calcul et de la possibilité de représenter graphiquement les résultats des calculs (Rodriguez-Galiano *et al.*, 2015).

De plus, leur utilisation s'avère très précieuse lorsque les techniques statistiques traditionnelles comme l'analyse discriminante ou la régression logistique atteignent leurs limites. En particulier, lorsque l'espace caractéristique est complexe. C'est-à-dire lorsque la dimensionnalité de cet espace est élevée et que la relation entre les dépôts ciblés et les fonctions de preuves d'entrées devrait être non linéaire. Ou encore, lorsque les jeux de données d'entrée devraient avoir des distributions statistiques différentes. Parmi ces catégories de classifieurs automatiques de plus en plus utilisés, figurent les arbres de décision. L'avantage de l'utilisation d'un algorithme basé sur une arborescence hiérarchique pour prendre des décisions de classification est que cette arborescence est transparente, ce qui en comparaison avec les RNA, est plus facile à interpréter (Rodriguez-Galiano *et al.*, 2015).

Sur la base de cette revue littéraire et compte tenu de la dimensionnalité spatiale élevée de notre zone d'étude et du manque de données d'exploration acquises sur les dépôts de coltan, c'est l'approche par apprentissage automatique non paramétrique qui a été préférée pour modéliser et cartographier de façon prédictive, la prospectivité de nouveaux dépôts. Il s'agira d'utiliser la puissance prédictive des algorithmes SVM, KNN et DT, pour parvenir à réaliser des classifications par apprentissage automatique, dans le cadre de cette étude de modélisation prédictive minérale.

2.5 Techniques de classification par apprentissage automatique

Ce paragraphe met en lumière les principes généraux des classifieurs automatiques et des fonctions utilisés pour entraîner des modèles (*Mdl*) sur un jeu de données d'apprentissage.

Présentation des Machines de support vectoriel (SVM) : Les classifieurs SVM effectuent une classification linéaire en trouvant un hyperplan qui maximise la marge entre deux classes distinctes, pour un ensemble fini de modèles d'apprentissage. Il s'agit de rechercher une règle de décision basée sur une séparation par un hyperplan de marge

optimale appelé vecteur de support (Sain *et al.*, 2006). La figure 2 présente de façon schématique un hyperplan de séparation de données (+,-) et trois vecteurs de support représentés dans un espace de fonction séparable linéairement.

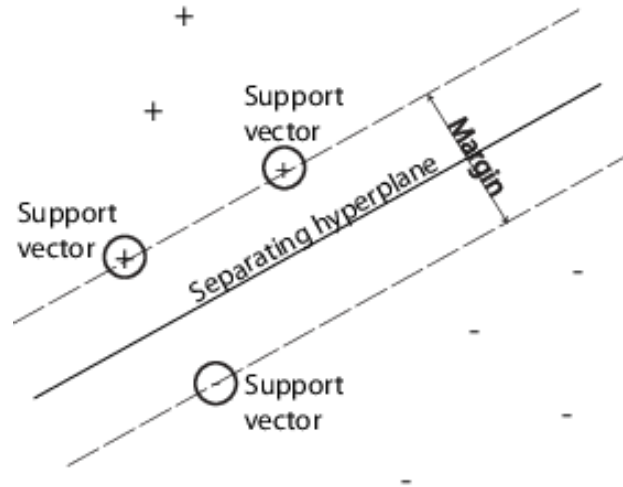


Figure 2 : « Tiré de MathWorks (1994) » Représentation schématique d'un hyperplan, de trois vecteurs de support et d'une surface de séparation de données (+,-).

Les modèles SVM linéaires sont généralement utilisés pour des cas d'apprentissage à deux classes. Le modèle SVM linéaire est un modèle de régression vectorielle à support, qui s'adapte à un modèle de régression linéaire, en supposant que les données d'apprentissage sont linéairement séparables. C'est-à-dire, il existe parmi une infinité de possibilités, un hyperplan à marge maximale capable de séparer les données sans erreurs. Dans ce cas, c'est la fonction « *fitlinear* » définie suivant l'équation suivante :

$$Mdl = \text{fitlinear}(X,Y) \quad (\text{Éq.1})$$

qui est utilisée pour former des modèles de classification linéaire pour l'apprentissage à deux classes (binaire) avec des données prédictives de grande dimension, complètes ou clairsemées. L'équation (Éq.1) renvoie un objet de modèle de classification linéaire formé qui contient les résultats de l'ajustement d'une machine à vecteur de support binaire aux prédictors X et aux étiquettes de classe Y . Les étiquettes de classe Y en sortie sont spécifiées sous forme de tableau de variables catégorielles, de caractères ou de chaînes de

caractères, de vecteurs logiques ou numériques ou sous forme de tableau de cellules de vecteurs de caractères. Les données de prédicteurs X sont spécifiées sous forme de matrice complète ou clairsemée. Les modèles de classification linéaire disponibles incluent les machines à vecteurs de support régularisés (SVM). *Fitlinear*, minimise la fonction objective en utilisant des techniques qui réduisent le temps de calcul (par exemple, la descente de gradient stochastique). Le modèle de classification linéaire formé est renvoyé comme modèle objet classifié ou comme modèle objet à validation croisée.

La prédiction des classes cibles par un modèle SVM linéaire entraîné, est effectuée à l'aide de la fonction suivante :

$$Label = predict(Mdl, X). \quad (\text{Éq. 2}).$$

« *Label* » contient des étiquettes de classe pour chaque niveau (ou force) de vecteurs régularisés dans *Mdl*. Cette fonction permet de renvoyer des étiquettes de classe prédites pour chaque observation dans les données de prédicteur X sur la base du modèle de classification linéaire binaire entraîné.

Les modèles SVM à noyaux binaires sont utilisés pour des cas d'apprentissage multiclasse non linéaire. Dans ce cas, la fonction « *fitkernel* » est définie selon l'équation suivante :

$$Mdl = fitkernel(X, Y) \quad (\text{Éq.3})$$

Les données d'un prédicteur X sont également spécifiées dans une matrice numérique $n \times p$, où n est le nombre d'observations et p est le nombre de prédicteurs. La longueur de " Y " et le nombre d'observations dans X doivent être égaux. Les étiquettes de classe Y sont également spécifiées sous forme de tableau de variables catégorielles, de caractères ou de chaînes de caractères, de vecteurs logiques ou numériques ou sous forme de tableau de cellules de vecteurs de caractères. *Fitkernel*, utilise une expansion aléatoire de fonctionnalités et entraîne ou valide de manière croisée un modèle de classification binaire à noyau gaussien pour la classification non linéaire. La fonction est plus pratique pour les applications gourmandes en données portant sur de grands ensembles d'apprentissage, mais elle peut également être appliquée sur des ensembles de données plus petits peu gourmands en mémoire. *Fitkernel* met en relation les données d'un espace de faible

dimension dans un espace de haute dimension, puis ajuste un modèle linéaire dans l'espace de haute dimension en minimisant la fonction objective régularisée. L'obtention du modèle linéaire dans l'espace à haute dimension équivaut à appliquer un noyau gaussien au modèle dans l'espace à basse dimension. Les modèles de classification linéaire disponibles incluent la machine à vecteur de support régularisé (SVM). Le modèle de classification à noyau formé est renvoyé également comme modèle objet ou comme modèle objet à validation croisée.

La prédiction des classes cibles par un modèle SVM linéaire entraîné, est également effectuée à l'aide de la fonction de l'équation (Éq. 2) et renvoie des étiquettes de classe prédites pour chaque observation dans les données de prédicteur X sur la base du modèle de classification binaire à noyau gaussien Mdl . Les « *Label* » ou étiquettes de classe prédites sont renvoyé(e)s sous forme de tableau de variables catégorielles, de caractères ou de chaînes de caractères, de matrice logique ou numérique ou de tableau de cellules (vecteurs de caractères). L'étiquette a " n " lignes (où n est le nombre d'observations dans X) a le même type de données que les étiquettes de classe observées (Y) utilisées pour former le Mdl .

Généralement, deux approches sont utilisées pour entraîner des modèles SVM sur un jeu de données d'apprentissage. La première approche *One vs One* (un contre un) tente de former un classifieur distinct pour chaque paire d'étiquettes différentes, en créant des « voteurs », où chaque voteur V_{ij} détermine si une entrée x a plus de chances d'appartenir à une catégorie i ou à une catégorie j . Un voteur V_{ij} est un SVM qui s'entraîne sur les données de catégorie i et j uniquement. En somme, classer une nouvelle entrée consiste à utiliser les éléments d'une classe contre toutes les autres, afin de retourner la catégorie qui aura remporté le plus de duels. La seconde est l'approche *One vs All* (un contre tous). Cette approche tente de classer toutes les classes en créant une classe binaire pour chaque classe, afin d'évaluer la probabilité d'appartenance d'un nouveau membre à l'une des classes binaires. En somme, classer une nouvelle entrée, consiste à entraîner un SVM sur chacun des couples de classes, puis à désigner la classe gagnante. L'opération est faite soit par un vote majoritaire, soit en post-traitant les résultats grâce à une estimation de probabilité a

priori. Ce qui permet d'évaluer à quelle catégorie la nouvelle entrée est la plus probable d'appartenir.

Présentation des arbres de décision (DT) uniques:

Les classificateurs DT disponibles dans le logiciel *Matlab*, effectuent une classification en utilisant la fonction «*fitctree*» pour représenter un ensemble de restrictions ou de conditions organisées hiérarchiquement. La fonction *fitctree* est utilisée pour ajuster un arbre de décision binaire dans les cas de classification multiclasse. Les conditions ou règles de décision sont appliquées successivement d'une racine à un nœud terminal ou à une feuille de l'arbre (Breiman, 1996). L'arbre est utilisé pour représenter de façon hiérarchique la structure des données sous forme de séquences de décisions (ou tests), en vue de la prédiction d'un résultat ou d'une classe. L'équation suivante :

$$Mdl = \text{fitctree}(Tbl, Y) \quad (\text{Éq. 4}),$$

renvoie un arbre de décision de classification binaire ajusté et basé sur les variables d'entrées contenues dans une table *Tbl*. La sortie est contenue dans le vecteur *Y*. Les observations utilisées pour former le modèle (*Mdl*) sont spécifiées sous forme de tableau (*Tbl*). Chaque ligne du *Tbl* correspond à une observation et chaque colonne correspond à une variable prédictive. Facultativement, "*Tbl*" peut contenir une colonne supplémentaire pour la variable de réponse.

Les arbres classifiés renvoyés sont des modèles d'arbres objets, car l'utilisation des options «*CrossVal*» (validation croisée) ou «*Kfold*» (K- sous-ensemble) aboutit à un modèle d'arbre de classe partitionné. Par conséquent, le résultat ne peut être utilisé pour une prédiction. Ce type d'arbre n'a pas de méthode de prédiction. Chaque méthode «*Kfold*» utilise des modèles formés sur les observations d'un sous-ensemble pour prédire la réponse pour des observations en dehors de ce sous-ensemble. Dans ce cas c'est la fonction «*KfoldPredict*» qui est utilisée pour prédire les réponses (ou étiquettes) ou «*Label*» des observations non utilisées pour la formation. L'équation suivante :

$$Label = \text{kfoldPredict}(obj) \quad (\text{Éq. 5})$$

permet de renvoyer les étiquettes de classe prédites par l'arbre objet « *obj* » suivant une classification à validation croisée (*K-fold*) appliquée. Pour chaque *fold* (ou sous-ensemble), *kfoldPredict* prédit les étiquettes de classe pour les observations du sous-ensemble K considéré.

Les arbres de décisions utilisent un critère d'optimisation. Le critère d'optimisation est l'homogénéité des descendants par rapport à la variable cible. La variable qui est testée dans un nœud sera celle qui maximise cette homogénéité. Le processus s'arrête quand les éléments d'un nœud ont la même valeur pour la variable cible (critère d'homogénéité). Chaque nœud est scindé (opération appelée *split*) de sorte à donner naissance à plusieurs nœuds descendants. Les arbres recherchent le meilleur *split* pour maximiser la pureté des arbres résultants, sur la base de nombreuses approximations. Il existe plusieurs règles ou critères de divisions qui peuvent être utilisé(e)s. On peut citer entre autres, l'indice de diversité Gini, le rapport de gain , le chi carré (Rodriguez-Galiano *et al.*, 2015) et la réduction maximale de déviance ou la règle de *Twoing*.

Les arbres de classification divisent les nœuds en fonction d'une impureté ou d'une erreur de nœud. L'erreur de nœud correspond à la fraction de classes mal classées au niveau d'un nœud.

Si j est la classe avec le plus grand nombre d'échantillons d'apprentissage sur un nœud, l'erreur de nœud est :

$$1 - p(j). \quad (\text{Éq. 6})$$

L'Indice de diversité de Gini d'un nœud est estimé selon le terme :

$$1 - \sum_i p^2(i) \quad (\text{Éq. 7})$$

Avec, la somme des nœuds des classes i , et $p(i)$ la fraction observée de classes de classe i qui atteignent le nœud. Un nœud avec une seule classe (un nœud pur) présente un indice de Gini de 0 ou une valeur positive.

Le paramètre de déviance maximale est estimé suivant le terme suivant :

$$- \sum_i p(i) \log p(i) \quad (\text{Éq. 8})$$

Avec, $p(i)$ la probabilité observée des classes de classe i qui atteignent le nœud. Un nœud pur présente une déviance de 0 ou une valeur positive.

La Règle de Twoing n'est pas une mesure de pureté d'un nœud, mais une mesure différente pour décider de la division d'un nœud. Soit $L(i)$ la fraction des membres de la classe i dans le nœud-enfant de gauche après une scission, et $R(i)$ la fraction des membres de la classe i dans le nœud-enfant de droite après une scission. Le mode de choix du critère de partage à maximiser se fait selon le terme suivant :

$$P(L)P(R)(\sum_i |L(i) - R(i)|^2) \quad (\text{Éq. 9}).$$

Où $P(L)$ et $P(R)$ sont les fractions d'observations qui se séparent à gauche et à droite respectivement.

Si l'expression est grande, la scission rend chaque nœud-enfant plus pur. De même, si l'expression est petite, la division a rendu chaque nœud-enfant similaire, par conséquent, similaire au nœud-parent, de sorte que la scission n'a pas augmenté la pureté du nœud.

Présentation des ensembles d'arbres de décisions :

Des modèles d'arbres de décision beaucoup plus complexes peuvent être utilisés dans cette étude. Il s'agit de modèles d'arbres de décision assemblés selon plusieurs méthodes. Cette approche permet de combiner un ensemble de modèles d'apprenants faibles suivant différentes méthodes. La prédiction de réponse de l'ensemble pour une classification de nouvelles données se fait en agrégeant les prédictions d'apprenants faibles. L'équation suivante est utilisée pour ajuster un ensemble d'apprenants pour une classification multiclasse:

$$Mdl = fitensemble(Tbl, Y) \quad (\text{Éq. 10})$$

La fonction traite toutes les variables du tableau Tbl comme des variables prédictives. Y est le tableau des étiquettes de classe. Les observations utilisées pour former le modèle sont spécifiées sous forme de tableau. Chaque ligne du Tbl correspond à une observation et chaque colonne correspond à une variable prédictive. Facultativement, Tbl peut contenir une colonne supplémentaire pour la variable de réponse. Les données de réponse sont spécifiées sous forme de tableau de variables catégorielles, de caractères ou de chaînes de caractères, de vecteur logique ou numérique ou de tableau de cellules de vecteurs de

caractères. Chaque entrée dans Y est la réponse ou l'étiquette de l'observation dans la ligne correspondante du Tbl . La longueur de " Y " et le nombre de lignes du Tbl doivent être égaux. Les modèles d'ensembles de classification croissent par rééchantillonnage et combinent un ensemble de modèles d'apprenants faibles entraînés à partir de données sur lesquelles ces différents apprenants ont été formés. Il peut prédire la réponse de l'ensemble sur de nouvelles données en agrégeant les prédictions de ses apprenants faibles.

Les techniques d'assemblage couramment utilisées produisent des ensembles d'arbres renforcés (*boosted trees*) ou ensachés (*bagged trees*). Ces techniques s'appuient sur une combinaison de plusieurs arbres de décision, afin d'améliorer leur performance prédictive dans les cas de classification multiple et complexe. Il s'agit entre autres des méthodes *Bagging* (ou ensachage ou encore appelée méthode *Random Forest* (Forêt aléatoire)) et *AdaBoost* (*Adaptive boosting* ou renforcement adaptatif) introduites par Freund (2009). Les modèles d'ensemble formés, retournent des modèles objets « *ClassificationBaggedEnsemble* » avec la méthode « *Bag* » ou « *ClassificationEnsemble* » pour toute autre méthode d'agrégation d'ensemble utilisé, en notamment dans cette étude, la méthode *AdaBoost*.

Méthode RF

La classification automatique par la méthode de *Bagging* ou forêt aléatoire (RF) est une technique de régression qui combine les performances de nombreux algorithmes d'arbres de décision, pour classer ou prédire la valeur d'une variable (Breiman, 1996; Rodriguez-Galiano *et al.*, 2015). Un modèle RF est obtenu avec la fonction *TreeBagger* et défini suivant l'équation suivante :

$$Mdl = TreeBagger(NumTrees, Tbl, Y) \text{ (Éq. 11)}$$

Celle-ci qui retourne un ensemble d'arbres de classification « *NumTrees* » en utilisant les variables prédictives du tableau Tbl et les étiquettes de classe du vecteur Y . Y est un tableau de données de réponse. Les éléments de Y correspondent aux rangées de la Tbl . Pour la classification Y est l'ensemble des véritables étiquettes de classe. Les étiquettes peuvent être n'importe quelle variable de regroupement, c'est-à-dire un vecteur numérique ou logique, une matrice de caractères, un tableau de chaînes de caractères, un tableau de

cellules de vecteurs de caractères ou un vecteur catégoriel. *TreeBagger* convertit les étiquettes en un tableau de cellules de vecteurs de caractères.

La fonction met en sac un ensemble d'arbres de décision pour la classification. L'ensachage signifie agrégation ou *bootstrap*. Chaque arbre de l'ensemble est cultivé sur une réplique *bootstrap* dessinée indépendamment des données d'entrée. Les observations non incluses dans cette réplique sont « hors du sac » pour cet arbre.

La fonction *TreeBagger* s'appuie sur la fonctionnalité « *ClassificationTree* » pour cultiver des arbres individuels. En particulier, « *ClassificationTree* » accepte un nombre d'entités sélectionnées au hasard pour chaque fraction de décision comme argument d'entrée facultatif. Autrement dit, *TreeBagger* implémente l'algorithme de la forêt aléatoire (RF) (Breiman, 2001). Les arbres de décision individuels ont tendance à s'adapter. Les arbres de décision agrégés par *bootstrap* (ensachés) combinent les résultats de nombreux arbres de décision, ce qui réduit les effets du surajustement et améliore la généralisation. *TreeBagger* développe les arbres de décision dans l'ensemble à l'aide d'échantillons *bootstrap* issus des données d'apprentissage.

Cette technique de rééchantillonnage aléatoire de l'ensemble des données d'origine avec remplacement, c'est-à-dire, sans suppression des données sélectionnées dans l'échantillon d'entrée pour générer le sous-ensemble suivant (Breiman, 2001; Rodriguez-Galiano *et al.*, 2015) confère une plus grande stabilité et robustesse au modèle de sortie, face à de légères variations dans les données d'entrée, tout en augmentant la précision de la prévision (Breiman, 1996). Une autre caractéristique intéressante selon Rodriguez-Galiano *et al.*, (2015), est que les arbres d'un classificateur RF poussent sans élagage, ce qui les rend légers, du point de vue du calcul. De plus, les échantillons qui ne sont pas sélectionnés pour l'apprentissage du $k^{\text{ième}}$ arbre, dans le processus d'ensachage, font partie d'un autre sous-ensemble appelé *out-of-bag* (*oob*). Le $k^{\text{ième}}$ arbre peut alors utiliser ces éléments (*oob*) pour évaluer ses performances (Peters *et al.*, 2007).

La prédiction de réponse d'un modèle d'ensemble d'arbre de décision ensachés est effectuée à l'aide de l'équation suivante :

$$Y_{fit} = predict(B, X) \quad (\text{Éq. 12})$$

La fonction *predict* renvoie un vecteur de réponses prédites pour les données de prédicteur dans un tableau ou une matrice *X*, basé(e) sur l'ensemble d'arbres de décision ensachés *B*. *Yfit* est un tableau de cellules de vecteurs de caractères pour une classification. Par défaut, *predict* prend un vote moyen démocratique (non pondéré) de tous les arbres de l'ensemble. *B* est un modèle objet *TreeBagger* entraîné, c'est-à-dire un modèle renvoyé par *TreeBagger*. *X* est un tableau ou une matrice de données prédictives utilisées pour générer des réponses. Les lignes représentent les observations et les colonnes représentent les variables. Lorsque *X* est une matrice numérique, les variables composant les colonnes de *X* doivent avoir le même ordre que les variables prédictives qui ont formé *B*.

La figure 3 présente de façon schématique un organigramme type, d'une classification utilisant un algorithme RF.

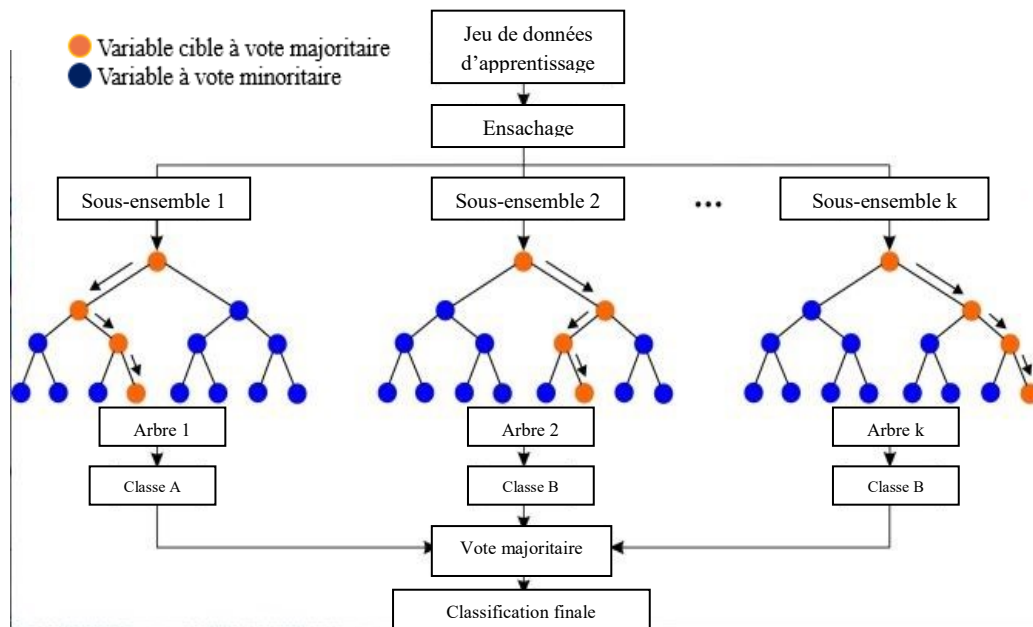


Figure 3 : « Modifié de Sun et al. (2019, p. 34) » Organigramme d'un arbre de classification utilisant un algorithme RF.

Présentation des K voisins les plus proches (KNN) :

K voisins les plus proches est un algorithme simple développé par Fix et Hodges en 1951 (Silverman *et al.*, 2006). Un modèle (*Mdl*) KNN est formé à l'aide de l'équation suivante :

$$Mdl = fitcknn(Tbl, Y) \quad (\text{Éq. 13})$$

L'équation renvoie un modèle de classification du plus proche voisin basé sur les variables prédictives de la table *Tbl* et du tableau de réponses *Y*. La longueur de la variable de réponse et le nombre de lignes dans *Tbl* doivent être égaux. Les observations utilisées pour former le modèle sont spécifiées sous forme de tableau. Chaque ligne du *Tbl* correspond à une observation et chaque colonne correspond à une variable prédictive. Facultativement, *Tbl* peut contenir une colonne supplémentaire pour la variable de réponse.

Le modèle de classification (k-voisin) formé est renvoyé comme modèle objet « *ClassificationKNN* » ou modèle objet à validation croisée « *ClassificationPartitionedModel* » lorsque l'option « Kfold » ou « CrossVal » est utilisée.

La prédiction de réponse (ou étiquettes) d'un modèle KNN entraîné s'effectue à l'aide de l'équation suivante :

$$label = predict(Mdl, X) \quad (\text{Éq. 14})$$

L'équation renvoie un vecteur d'étiquettes de classe prédites pour les données de prédicteur dans une table ou une matrice *X*, sur la base du modèle de classification *Mdl* (des k plus proches voisins) formé.

Le principe consiste à stocker tous les cas disponibles et à classer les nouveaux cas en fonction d'une mesure de similarité. Il s'agit de classer des points d'une requête en fonction de leur distance par rapport à d'autres points (en l'occurrence les plus proches voisins) dans un jeu de données d'apprentissage. Cette méthode est un moyen simple, mais efficace de classer de nouveaux points. C'est une technique des plus simples et des plus intuitives

largement utilisée dans le domaine de la discrimination statistique. Il s'agit d'une méthode non paramétrique, dans laquelle une nouvelle observation est placée dans une classe d'observation parmi l'ensemble d'apprentissage, le plus proche de la nouvelle observation, en ce qui concerne les covariables utilisées (Hechenbichler *et al.*, 2004). Dans cette méthode, une distance (euclidienne ou autre) est assignée entre tous les points d'un jeu de données. Les K plus proches voisins (K étant le nombre de voisins) sont ensuite trouvés par analyse ou par tri de leur matrice de distance. Les K plus proches voisins sont par la suite analysés, pour déterminer quelle étiquette de classe est la plus commune parmi l'ensemble « voté » le plus proche. L'étiquette de classe la plus courante est alors affectée au point de données en cours d'analyse.

La figure 4 illustre le mode de fonctionnement d'une classification utilisant les KNN.

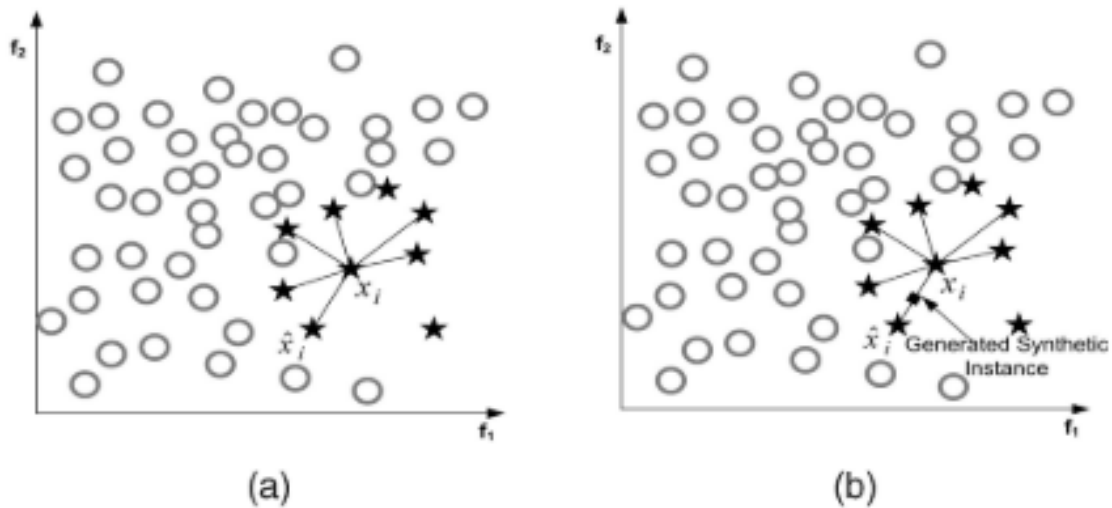


Figure 4 : « Tiré de He et al. (2009, p.1267) » (a) Exemple des K-voisins les plus proches pour une observation x_i considéré ($K = 6$). (b) Création de données basée sur la distance euclidienne.

Une autre variante de modèle KNN (KNN « *weighted* » ou « pondéré ») peut être utilisée. Cela consiste à affecter un « poids » aux différents voisins, de sorte à influencer le classement d'un échantillon en fonction de la distance à ses proches voisins (Balabin *et al.*, 2011). Cet algorithme beaucoup plus flexible introduit un système de pondération pour les voisins les plus proches en fonction de leur similarité avec une nouvelle observation à

classer, du fait que le vote des voisins les plus proches est équivalent au mode de distribution de probabilité de la classe (Hechenbichler *et al.*, 2004).

2.6 Outils d'aide à la décision pour la recherche minière

L'objectif final du géologue minier est d'estimer de nouvelles ressources économiques au sein de vastes étendues géographiques. Les résultats d'une étude de cartographie prédictive minérale sont des cartes de potentialités minérales utilisables pour prospecter de nouvelles cibles potentiellement économiques. L'identification de ressources minérales disponibles nécessite de prospecter en surface et en profondeur ces nouvelles cibles potentielles (parfois sur de vastes étendues géographiques). Des outils cartographiques interactifs connectés sont de plus en plus utilisés dans le cadre de la préparation d'opérations de levés plus approfondis.

L'utilisation et le traitement de données géologiques et géophysiques régionales complémentaires améliorent les connaissances acquises sur la minéralisation en présence et servent de guide au prospecteur pour la planification de futurs travaux de terrain (levés géophysiques, levés géologiques, levés géochimiques, etc.). Des données cartographiques administratives (découpage territorial, cadastre minier, etc.) et/ou d'occupation du sol (végétation, cours d'eau, routes, pistes, etc.) guident le prospecteur sur la disponibilité ou l'accessibilité des sites. L'affichage de cartes de potentialités minérales superposables (au sein d'une seule plateforme interactive) à ces différentes cartes thématiques donne accès à un ensemble d'informations spatiales géoréférencées. Des informations stratégiques sont alors utilisables pour orienter, de façon efficiente, des travaux de prospections avancées. Ce qui peut contribuer à réduire considérablement la durée et les coûts liés à la compilation et au traitement de données pluridisciplinaires, dispersées et souvent volumineuses.

Plusieurs approches méthodologiques peuvent être utilisées pour diffuser de l'information spatiale en matière de recherche minière. Plusieurs travaux (Weissman *et al.*, 2009; Song *et al.*, 2010; Wu *et al.*, 2015; Lanucara *et al.*, 2017) ont porté sur des méthodes utilisables pour développer un outil d'aide à la décision, capable de gérer, partager et visualiser l'information de géologie minière. Ces approches utilisent généralement un SIG *Web*, pour diffuser et gérer de façon analytique des données géospatiales liées à un environnement

naturel. Le choix des outils de développement et d'applications personnalisées est intrinsèquement lié aux spécificités technologiques du projet. Il reste toujours à l'appréciation du développeur (Chen *et al.*, 2014a). Il est régi selon les exigences d'interopérabilité des données, la pérennité, la performance, la sécurité, la flexibilité du système ou selon la complexité des tâches de géotraitement à exécuter.

L'une des approches consiste à construire une architecture SIG *Web*, capable de diffuser de façon interopérable et au plus près de la source, des données géospatiales stockées dans un système de gestion de base de données à référence spatiale (SGBRS). Ces solutions peuvent être développées pour garantir un accès à une information géospatiale de façon démocratique (sans logiciel SIG nécessaire). Les solutions dites *Open Source* (libres et ouvertes) figurent au nombre de celles-ci. Elles sont peu coûteuses et utilisent des outils conviviaux, libres et ouverts pour diffuser un catalogue de données géospatiales, suivant des normes internationales, telles que celles de l'*Open geospatial consortium* (OGC). Ce type d'architecture garantit l'interopérabilité, la pérennité, la flexibilité et la performance du système (Touzet *et al.*, 2017). L'approche consiste à concevoir un SGBRS, à programmer un serveur cartographique et une interface cartographique *Web*, pour réaliser finalement une application cartographique ou un géoportail accessible depuis plusieurs plateformes (téléphone intelligent, tablettes ou ordinateurs). Par conséquent, cette application peut être utilisable comme outil d'aide à la décision dans le cadre du ciblage rapide de nouvelles zones prospectives.

À titre d'exemple, les travaux de Gkatzoflias *et al.* (2013), portant sur le développement d'une infrastructure SIG *Web Open source*, ont permis, de diffuser un modèle de répartition spatiale des polluants et des sources d'émissions de la République de Chypre. Au niveau de l'architecture SIG *Web*, le développement de l'application a nécessité d'utiliser des outils logiciels *open source* (Mapserver, PostgreSQL et PostGIS), pour gérer des données géospatiales et des langages de programmation (HTML (*Hypertext Markup Language*), PHP (*Hypertext Preprocessor*) et JavaScript) pour gérer la communication entre les différents outils du système. La mise en place d'un tel système a favorisé la gestion de cartes vectorielles, matricielles, ainsi que la gestion de mosaïques de cartes avec une

résolution temporelle. En outre, l'architecture développée est peu coûteuse et accessible à des utilisateurs sans aucune connaissance préalable du champ SIG.

D'autre part et en matière de recherche minière, l'interopérabilité de l'information géospatiale est un critère nécessaire. Car, les SIG *Web* sont de plus en plus utilisés par des acteurs (gestionnaires et décideurs du secteur minier) pluridisciplinaires et multi localisés. À cet effet, cette dernière décennie, de nouvelles technologies de partage de l'information géospatiale ont été mises en place pour satisfaire au maximum des spécificités de différentes architectures SIG *Web* multi localisées.

La technique consiste à mettre en place une architecture intégrant des technologies distribuées. Il s'agit de mettre en réseau, différents SIG *Web* orientés objets, afin de lever des problèmes de sémantiques ou d'ontologies hétérogènes. Cela permet en outre, de résoudre les problèmes de mises à jour, liés aux sources d'informations géospatiales multi localisées. Cette technique est connue sous le nom de l'infonuagique ou « Informatique Nuagée » (ou encore *Cloud Computing*). Elle a de plus en plus été étudiée et appliquée dans plusieurs travaux récents (Evangelidis *et al.*, 2014; Ravikumar *et al.*, 2013; Song *et al.*, 2010; Weissman *et al.*, 2009). Il s'est agi d'implémenter dans une architecture SIG *Web*, des serveurs objets distribués qui fonctionnent en temps réels (Weihua, 2009; Balestro *et al.*, 2013; Zhang *et al.*, 2015). Dans ces cas, la diffusion de l'information n'est plus axée sur les données elles-mêmes, mais sur les métadonnées (Ji *et al.*, 2007). La stratégie est basée sur l'adoption de principes d'ontologies communes, fixant des règles de partage, par le biais d'entités graphiques standardisées et normalisées. Ce qui permet d'harmoniser les représentations visuelles et textuelles de différents modèles de données. Cette approche permet de garantir non seulement, l'interopérabilité des cartes, mais surtout celles de données géominières nationales et internationales normées.

A titre d'exemple, Lanucara *et al.* (2017) ont développé un modèle de données UML (*Unified Modeling Language*)(GeoSciML) et un langage de balisage géographique GML (*Geography Markup Language*) basés sur des standards WFS (*Web Feature Service*). Cette technologie permet de modéliser et d'harmoniser des données, suivant un schéma d'application simple GML (GeoSciML-Portayal). Ce schéma est basé également sur des standards WMS (*Web Map Service*) et SLD (*Styled Layer Descriptor*) autorisant un partage

interopérable et harmonisé des représentations visuelles. Cette technique a permis de réaliser deux modèles associés à des thésaurus géoscientifiques d'un système simple d'organisation de connaissances. Ce qui a permis d'enrichir de façon sémantique des données et des cartes. De plus, l'architecture propose une interface basée sur un schéma XML (*Exchange Markup Language*) compatible aux normes de l'OGC.

Ce schéma permet de surmonter les différences de langages graphiques vectoriels SVG (*Scalable Vector Graphics*), JavaScript ou des tables HTML, communément utilisés, soit pour le partage de données d'échantillons géologiques, soit pour le partage de données analytiques géochimiques et géochronologiques. Par conséquent, cette interface peut être utilisée pour le partage interopérable de mesures d'observations et d'échantillonnage, très utilisée en matière de recherche minière. Fort de ces avantages, cette technologie est largement intégrée dans plusieurs géoportails, dédiés à la recherche minière à travers le monde. Pour n'en citer que quelques exemples ; c'est le cas des géoportails australiens (*GeoScience Portal*) (Scott *et al.*, 2014) et italiens (GeoSciML-Portayal).

Dans le cadre de ce projet de recherche, il s'est agi de mettre en place dans un premier temps, un prototype de géoportail minier, dédié à la diffusion de l'information géominière complémentaire et utilisable pour guider des travaux de prospections avancés. À long terme, il s'agira de faire migrer le système en fonction des divers besoins et enjeux stratégiques, posés en matière de recherche et de partage de données géominières. À ce stade de l'étude, il s'est agi uniquement de développer un outil basé sur une architecture SIG *Web open source*, en vue de vulgariser au maximum, l'accès à l'information par un géoportail accessible depuis plusieurs plateformes mobiles connectées (ordinateur portable, téléphone intelligent et tablette). Car, dans un contexte de pays en voie de développement, l'accès aux téléphones intelligents est plus répandu que celui des ordinateurs portables ou de bureau.

C'est dans ce contexte, qu'une architecture SIG *Web open source* a été choisie, pour réaliser notre outil d'aide à la décision. La figure 5 présente une architecture SIG *Web Open Source* type, utilisable pour diffuser de l'information géospatiale et garantissant une bonne interopérabilité des données stockées dans un SGBRS.

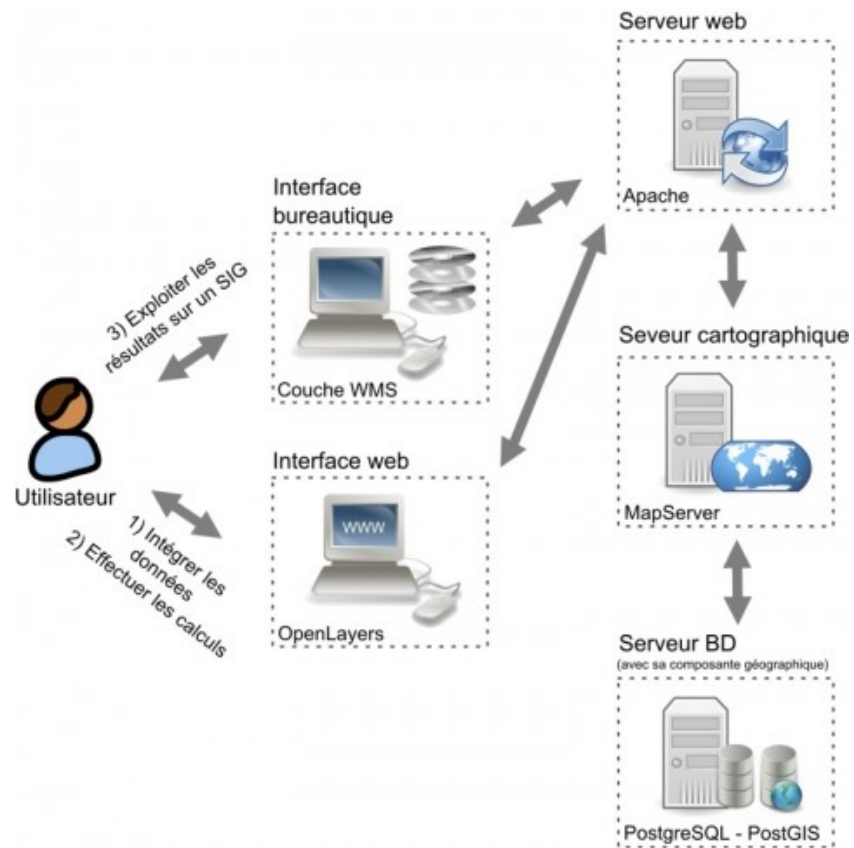


Figure 5: Architecture SIG Web formée avec des outils Open Source

CHAPITRE III. CADRE EXPÉRIMENTAL

PRÉSENTATION DE LA ZONE D'ÉTUDE

3.1 Sites d'études

Situation géographique

Les zones d'étude couvrent une superficie totale de 645,51 km² et se situent entre les latitudes 6° et 7° N et les longitudes 6° et 7° W. La partie nord, est localisée dans la région d'Issia au centre-ouest de la Côte d'Ivoire. La ville d'Issia est située à 384 km d'Abidjan, à 175 km de Yamoussoukro et à 49 km de Daloa, sa capitale régionale. Les sites de Bémadi, Makua et Brokoua situés au nord du Complexe ont été sélectionnés pour la disponibilité des données de puits effectués dans le cadre d'un projet d'exploration réalisé en 2015. La figure 6 situe ces trois (03) sites d'étude.

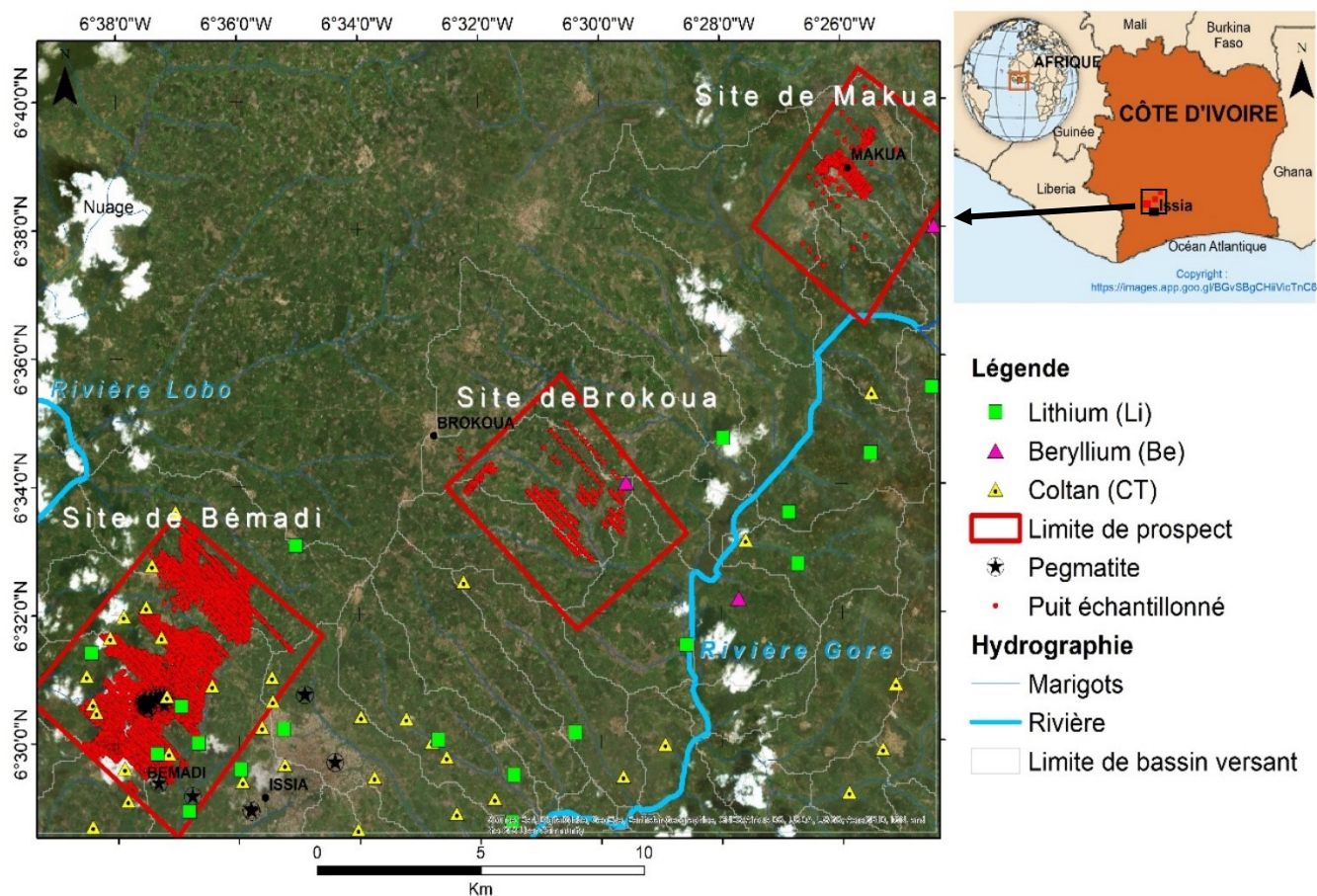


Figure 6 : Situation géographique de la partie nord de la zone d'étude

La partie centrale du Complexe géologique se situe au sud-ouest de la Côte d'Ivoire dans la région de Soubré et abrite deux nouveaux sites sélectionnés pour les besoins de l'étude. La ville de Soubré est située au sud-ouest du pays, à 400 km d'Abidjan. Les sites de Grand-Zatry et Oureyo (peu prospectés) présentés à la figure 7 ont fait l'objet de notre campagne de terrain en 2018. Ils ont été sélectionnés dans le but d'évaluer la pertinence des résultats de l'étude sur la partie sud du Complexe géologique.

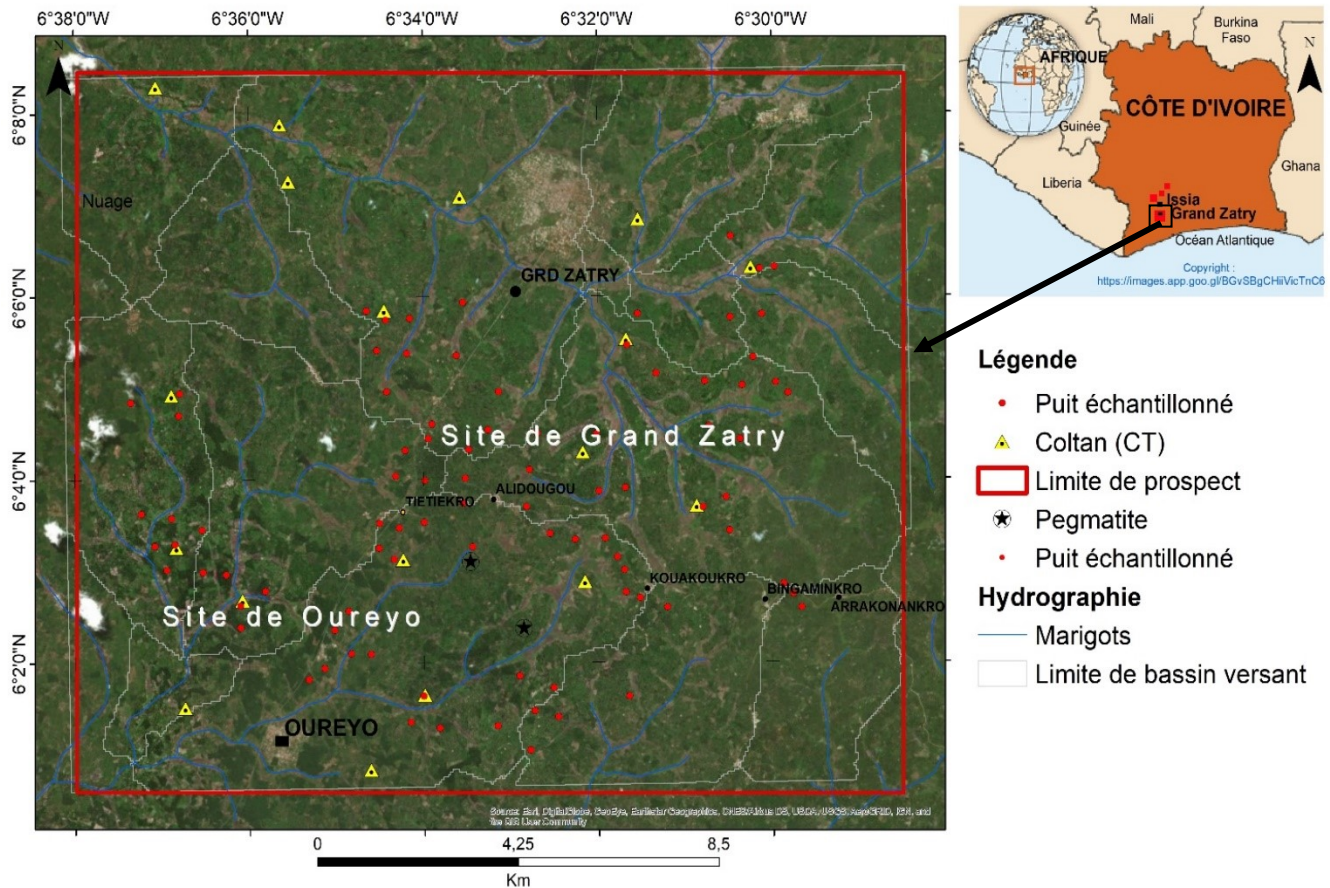


Figure 7 : Situation géographique de la partie sud de la zone d'étude

Toutes les formations géologiques, essentiellement composées de roches volcano-sédimentaires, sont drainées par les sous bassins versants de la Lobo. Ils sont représentés avec les lignes blanchâtres à la figure 7. Ces surfaces hydrographiques ont été sélectionnées pour l'étude, en raison de la présence avérée de pegmatites granitiques et/ou d'indices minéraux de coltan. Les affleurements rocheux de pegmatites granitiques minéralisés en coltan (représentés par les étoiles noires) et les indices alluvionnaires, colluvionnaires et éluvionnaires de coltan (CT) (représentés par les triangles jaunes) aux figures 6 et 7 ; de

béryllium (Be) et de lithium (Li) (représentés respectivement par les triangles roses et les carrés verts) à la figure 6, ont été mis en évidence par des travaux de prospections systématiques réalisés sur les flancs de collines, dans les marigots et les rivières de ces régions (Rombach, 1960; Cruys, 1965).

Climat, végétation et sol

Les secteurs d'études sont soumis à un climat subtropical humide, marqué par des températures annuelles allant de 25 °C à 30 °C. Le climat est caractérisé par un régime alterné de deux saisons pluvieuses et de deux saisons sèches avec des précipitations annuelles moyennes comprises entre 1200 et 1600 mm. La grande saison des pluies s'étend d'avril à juin, la petite saison sèche intermédiaire de juillet à mi-septembre, la petite saison des pluies, de la mi-septembre au mois de novembre et la grande saison sèche, de décembre à mars, avec des pointes de chaleur. Les saisons pluvieuses sont marquées par des précipitations fortes et régulières, tandis que les saisons sèches sont caractérisées par une forte sécheresse et de rares pluies. Toutefois, avec les effets du réchauffement climatique, l'on observe de plus en plus, une irrégularité de la pluviométrie d'une année à l'autre.

Les régions étudiées sont situées en domaine forestier tropical. La végétation des sites de Bémadi, Makua et Brokoua a été affectée par de nombreuses cultures pérennes et vivrières. Par conséquent, la forêt primaire de type équatorial a laissé place à une forêt secondaire. De même, au niveau des sites de Grand Zatry et Oureyo, la végétation est marquée par une forêt noire, dense et humide, qui fait place aujourd'hui à des lambeaux de forêts et d'immenses plantations de cultures pérennes traditionnelles ou industrielles. Sur l'ensemble de la zone d'étude, les zones d'affleurement des massifs rocheux sont généralement couvertes d'herbes basses. D'une façon générale, il est fréquent de trouver dans ces régions, que de grands arbres ne viennent pas jusqu'aux marigots, mais que ceux-ci soient bordés d'essences arbustives plus basses ou même seulement de hautes herbes. En revanche, certaines rivières (les plus importantes) sont bordées de forêts qui y plongent littéralement (Papon *et al.*, 1971).

Les sols du centre-ouest forestier de la Côte d'Ivoire sont essentiellement de type ferrallitique moyennement désaturé et évoluent progressivement vers le type ferrallitique faiblement désaturé en horizon B, compte tenu de la diminution progressive des

précipitations annuelles. Ils sont caractérisés par la présence d'un horizon riche en éléments grossiers (débris de cuirasse, gravillons ferrugineux, graviers et cailloux de quartz plus ou moins émoussés) (Perraud, 1971 *in* Allou, 2005 p.17). Les matériaux détritiques sont généralement constitués de dépôts argilo-sableux ou sablo-argileux de faible épaisseur (1 à 5 m) avec présence de nombreux graviers.

Contexte géomorphologique régional

Les sites de Grand Zatry et d'Oureyo présentent un relief caractérisé par de vastes plateaux que surmontent par endroits quelques élévations constituées de collines et de monts. Les monts Troquai situés dans la sous-préfecture d'Okrouyo, présentent trois sommets de formes ramassées dominant à 426 m, 444 m et à 445 m. Ces monts surplombent un paysage constitué de pénéplaines situées en moyenne à 162 m d'altitude. Également, il est possible d'observer une série de petites hauteurs alignées culminant de 216 m à 246 m à certains endroits.

Le relief de la région d'Issia est assez plat et entrecoupé par de nombreux basfonds. La zone est affectée par une évolution morphologique très poussée approchant beaucoup les caractères des pénéplaines.

Plusieurs dômes granitiques constitués de masses cristallines émoussées surplombent à certains endroits le paysage, à une hauteur de 100 m. Ces dômes sont des masses résiduelles du socle précambrien plus ancien et probablement plus élevé. Le relief est marqué par un ensemble de thalwegs et d'interfluves réduit à une surface de dénivellations peu importantes. Cette topographie remarquablement faible et régulière présente des formes assez basses et monotones, mollement taillées par de nombreux vallons à fonds plats et encadrés de pentes douces à profil largement concave.

Au plan morphologique, les régions nord et sud du Complexe, peuvent être regroupées en trois unités géomorphologiques d'âges différents.

La première correspond au niveau le plus récent et est constituée par des flats subhorizontaux de rivières et de marigots de largeurs variables. La seconde est constituée de vastes plateaux occupés par de larges interfluves s'étalant à une altitude moyenne de 200 m. Les zones d'interfluves ne dominent que peu les zones de flats avec des

dénivellations de l'ordre de 10 à 15 m maximum (Cruys, 1965). La troisième unité (située au nord du Complexe) est constituée de dômes granitiques et de collines surbaissées, dont la base se raccorde progressivement avec le niveau des interfluves.

Contexte hydrographique régional

Les sites de Bémadi, Brokoua et Makua appartiennent au grand bassin versant de la Lobo. Ce bassin s'étend entre les latitudes 6°05 et 7°50 N et les longitudes 5°35 et 7°38 W. La figure 8 présente la situation géographique de ce bassin versant. Le bassin est alimenté par des affluents et sous affluents qui épousent pour la plupart la direction des structures géologiques majeures, notamment les principaux systèmes de fracturation du socle (Allou, 2005). La rivière Lobo est l'un des principaux affluents rive gauche du fleuve Sassandra. La rivière évolue suivant une direction générale nord-sud, en dessinant de nombreux méandres jusqu'à son exutoire situé au niveau de Loboville (figure 8). Le réseau hydrographique est strictement contrôlé par des lignes structurales régionales en particulier par le principal système de diaclases du vieux soubassement cristallin de direction N 110° à 150° E, N 60° à 90° E et N 0° à 30° E (Cruys, 1965).

Les sites d'étude de Grand Zatry et d'Oureyo sont également caractérisés par un réseau hydrographique qui comprend principalement le fleuve Sassandra et la Lobo et plusieurs rivières et marigots.

De façon générale, le régime des rivières est essentiellement irrégulier. En période de pluies, les débits sont forts et les crues importantes. Mais les débits diminuent rapidement après les grandes pluies et la plupart des marigots tarissent plus ou moins vite pendant la saison sèche. La figure 8 situe et présente le grand bassin versant de Lobo.

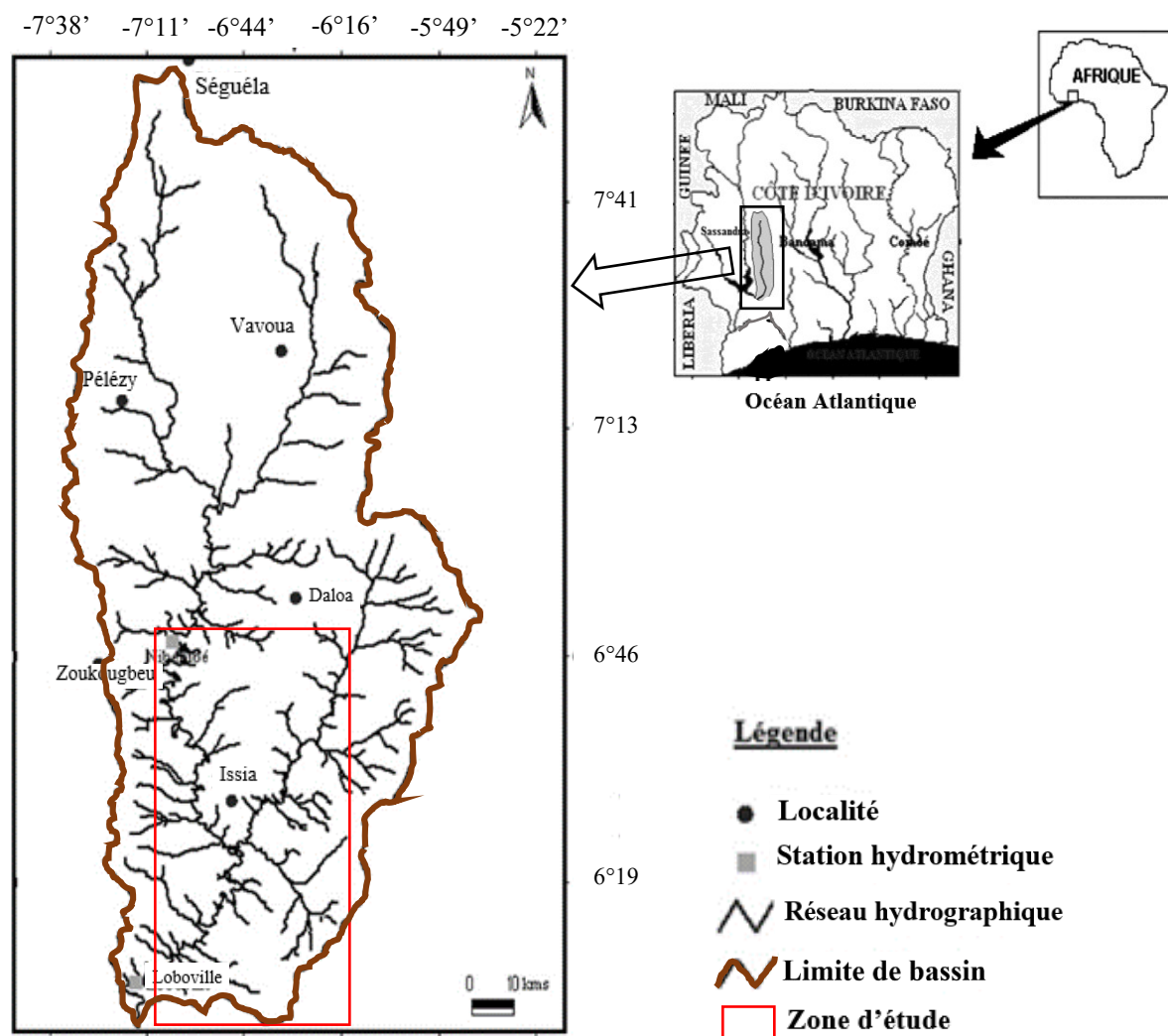


Figure 8 : « Modifié de Yao et al. (2016, p. 857) ». Localisation du bassin versant de la Lobo

Contexte géologique régional

Le Complexe géologique Hana-Lobo a été étudié par plusieurs auteurs (Meloux, 1959; Arnould, 1961; Papon, *et al.*, 1971). Il s'étend du Sud au Nord, entre les latitudes 4°55 et 6°40 N. La figure 9 présente une esquisse géologique du Complexe. Il s'agit d'une série géosynclinale épi à méso métamorphique plissée, limitée au sud-ouest par les formations Nigré et Mont Kedio et à l'Est par l'Unité volcano- sédimentaire de la Davo. La limite nord-ouest est marquée par la formation Buyo-Guessabo, la limite sud, par le littoral ivoirien (Océan Atlantique) et le centre est occupé par la formation de Zakoeoua, située au niveau de la localité d'Oureyo.

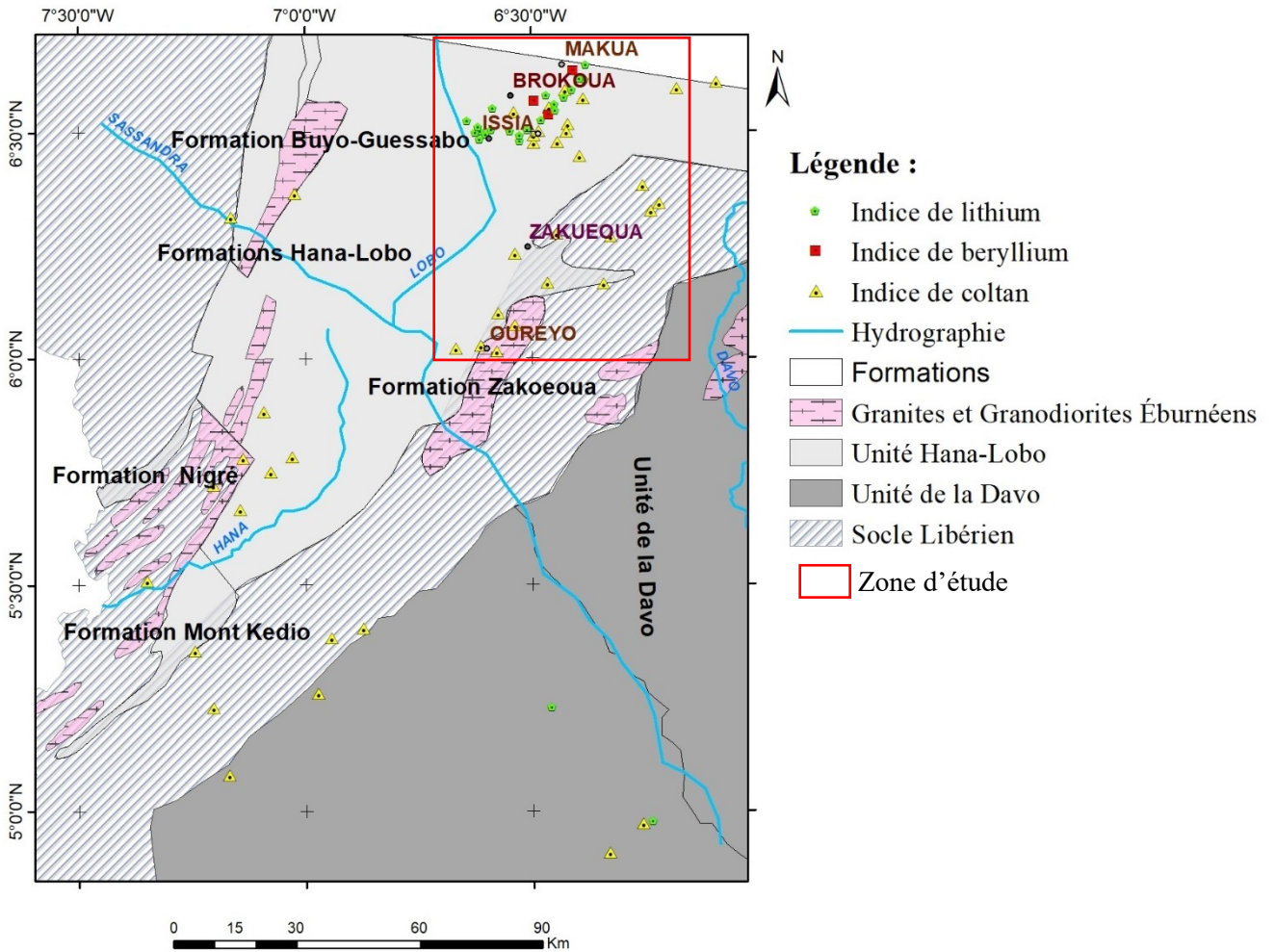
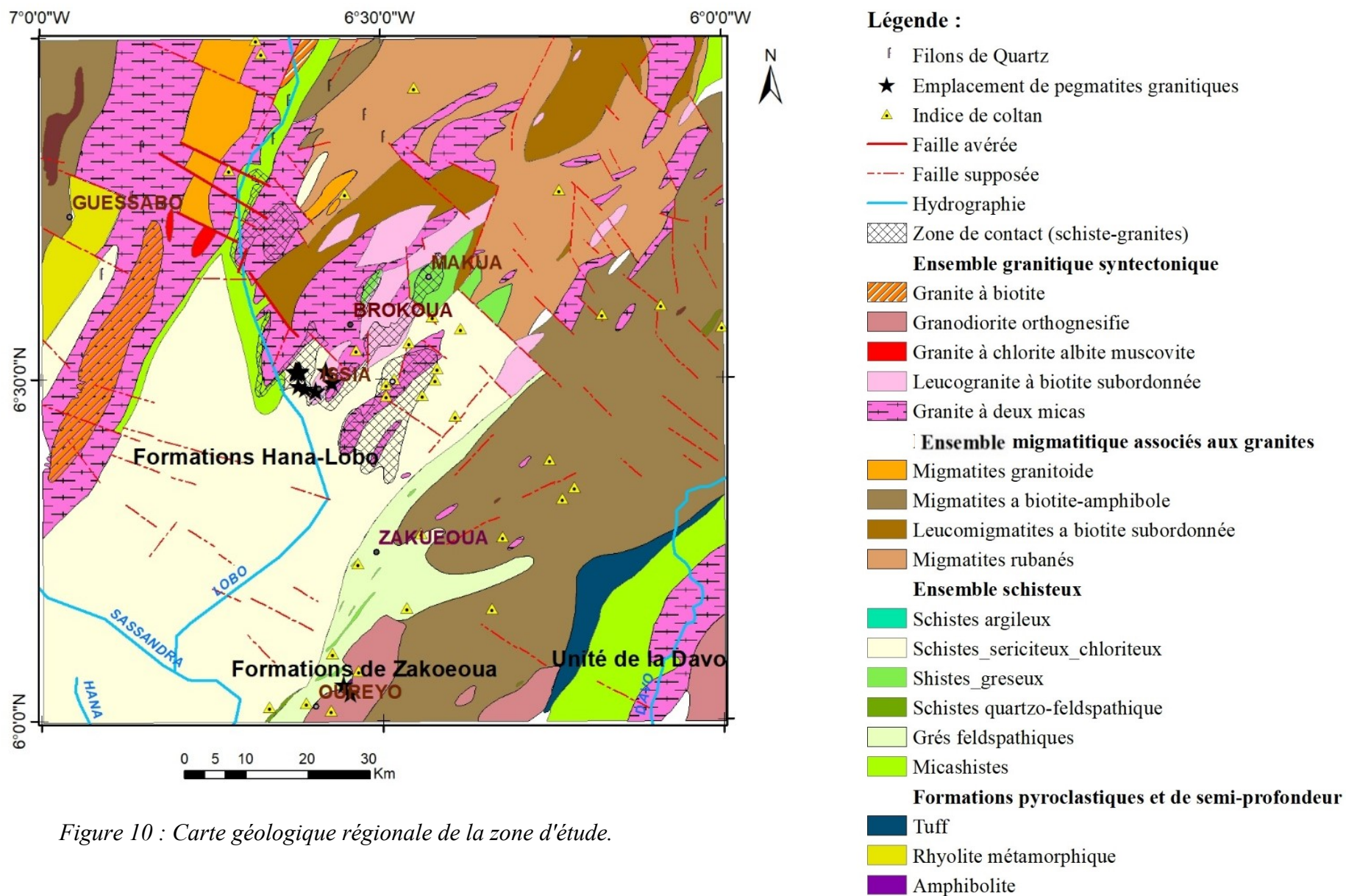


Figure 9 : « Modifié de Papon et al. (1971, p. 23) ». Esquisse du complexe géologique Hana-Lobo-Davo.

Au plan géologique, l'histoire du Complexe est marquée par le jeu d'une tectonique qui a opéré des décrochements et des remontées successives du socle libérien. Au niveau structural, plusieurs failles de direction ouest-nord-ouest tronçonnent la bordure ouest du Complexe et remontent le compartiment sud. La figure 10 présente une carte du contexte géologique de la région. La carte mentionne les grands accidents majeurs (failles, fractures) et les formations géologiques. La carte situe également les affleurements de pegmatites granitiques observés lors des anciens travaux de terrain, ainsi que les indices minéraux de coltan connus.

Au point de vue structural, une série de failles orientées NNW et EW affectent les granites et les granodiorites éburnéens au niveau des formations Mont Kedio, Nigré et Buyo-Guessabo. Au niveau lithologique, l'Unité géologique Hana –Lobo est constituée essentiellement de roches metabasites (roches vertes), de flyschoides (micaschistes, schistes) et de roches volcano-sédimentaires. La formation Hana-lobo située au nord du Complexe, héberge les dépôts de colombo-tantalite connus de la région. La partie sud de la zone d'étude, où plusieurs indicateurs minéraux de présence du minerai de colombo-tantalite ont été identifiés, appartient à la formation Zakoeoua. Les élévations granitiques de la région nord du complexe (région d'Issia) se sont développées à la suite d'un important métamorphisme de contact ayant affecté les schistes mitoyens. Les granites apparaissent sous forme de dômes et se prolongent vers le Nord-Est. La carte géologique régionale (figure 10) montre que la formation Hana-Lobo se compose de granites leucocrates homogènes, de granites calco-alcalins et de granites à deux micas. Les formations géologiques minéralisées des sites de Bémadi, Brokoua et Makua sont d'âges récents et du précambrien. Le socle précambrien est un domaine de granites métagénétiques intrudés de filons et de roches essentiellement schisteuses d'origine sédimentaire. Ces formations se composent de roches séricito-schisteuses, chlorito-schisteuses, de schistes gréseux, micacés et arkosiques et de micaschistes. Rarement visibles en affleurement, ces roches sont observées dans des puits de prospection à l'état altéré (argiles jaunes, brunes et trainées blanchâtres, rougeâtres ou violacées).



Actuellement, cette région, du fait d'une géomorphogenèse exosphérique assez prononcée, est dans sa presque totalité recouverte par des dépôts éluvionnaires, colluvionnaires et alluvionnaires qui occupent la quasi-totalité de la surface. Ce sont ces dépôts détritiques superficiels (principaux réservoirs des métaux lourds) qui font l'objet de notre étude. La figure 11 donne un aperçu d'une coupe schématique caractéristique des dépôts étudiés. Les gisements sont constitués de graviers minéralisés et sont généralement de faible amplitude. Ils se localisent près de la surface, sous une faible couverture végétale et en aplomb d'une couche de micaschistes ou de roches pegmatitiques peu altérées. Par endroit, une couche argileuse, sableuse, argilo-sableuse ou sablo-argileuse est rencontrée, avant d'atteindre le premier gravier. Les dépôts sont constitués le plus souvent de deux couches graveleuses de granulométries fines, moyennes ou grossières.

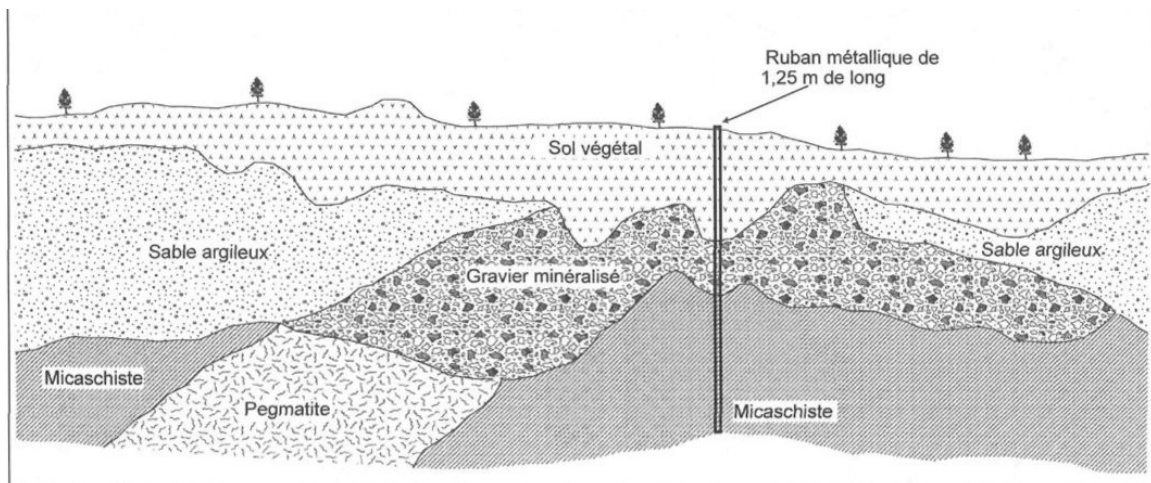


Figure 11 : « Tiré de Allou (2005, p.196) ». Architecture générale du manteau d'altération de la mine Étienne-Méguhé.

Sur le terrain, les coupes de puits réalisées en 2015, ont confirmé la présence effective de pegmatites peu altérées, de micaschistes et d'un ou de deux niveaux de graviers latéritiques de granulométries différentes. L'observation des profils de ces puits confirme aussi que les graviers recouvrant le socle précambrien sont mêlés dans une matrice argileuse à argilo-sableuse au niveau des éluvions et des colluvions. La matrice devient de plus en plus sableuse à l'approche des alluvions.

3.2 Données utilisées

Le tableau 1 présente les spécificités des données utilisées.

Tableau 1: Tableau de synthèse des données utilisées

Nature	Format	Source	Résolution spatiale	Résolution spectrale	Paramètre étudié	Utilisation
Minéralogie	Vectorel (points)	Projet d'exploration coltan (SODEMI)	25 m x 50 m		Teneur en concentré de coltan dans les graviers	Identification de cibles minéralisées
Hydrologie	Matriciel	ASTER (ASTGTM2_N06 W007_dem) du 17 octobre 2011	30 m		Transport, érosion et dépôt des minéraux de coltan (Surface d'accumulation de flux ; Amplitude ou direction de flux)	Discrimination de zones favorables au transport et au dépôt de minéraux lourds
Morphologie	Matriciel	ASTER (ASTGTM2_N06 W007_dem) du 17 octobre 2011	30 m		Transport, érosion et dépôt des minéraux de coltan (Degré de pente, Profil et courbure de pente, Élévation)	Identification de zones d'irrégularités topographiques favorables au piégeage de minéraux lourds
Structurale	Matriciel	Sentinel 2B MSI de niveau 1C du 24 décembre 2017	10, 20, 60 m	13 bandes de longueurs d'onde (443 à 2190 nm)	Discontinuité dans les pixels	Extraction de linéaments
		Sentinel 1A SAR IRW GRDH du 6 janvier 2018 (Polarisation VV/HH)	40 m pour 22 x 5 regards			
		Imagerie ESRI (Terra Color) SPOT	15 m 2,5 m	,	Apparence haute résolution des objets sur l'image	Validation des failles et fractures naturelles
Géophysique	Vectorel (polygones)	SODEMI	1/50 000 ; 1/ 200 000		Couverture SIG et disponibilités de données géophysiques (levés au sol ou aéroportés)	Accessibilité aux données de géophysique existantes ou à acquérir (magnétique, gravimétrique ou électrique) en vue d'études et travaux de terrain complémentaires
Géologie	Vectorel (polygones, polygones, points)	Carte géologique régionale (DGMG)	1/200 000		Surfaces lithologiques et localisation de Pegmatite granitiques, failles et fractures régionales	Localisation et compréhension du contexte géologique de minéralisation
Administrative minière	Vectorel (polygones, polygones, points)	DGMG/ BENTD	Échelle graphique		Couverture SIG, accessibilité aux zones prospectives prédites et responsabilité administrative/cadastrale de sites d'intérêts	Localisation administrative (sous-préfecture) et cadastrale (permis accordé ou non) des sites d'intérêts pour des travaux de terrain et études complémentaires

Les paragraphes suivants en détaillent les spécificités et motivent leurs utilisations dans le cadre de ce projet de recherche.

3.2.1 Données géologiques

Les données géologiques au format vectoriel (points, polygones et lignes) proviennent de la feuille géologique NB-29-XVIII à l'échelle 1/200 000. La carte a été éditée par l'institut Géographique de Côte d'Ivoire le 9 avril 1990. L'information géologique utilisée pour l'étude, est constituée des emplacements d'affleurements rocheux de pegmatites granitiques, des indices minéraux de coltan, de béryllium, de lithium connus, des formations granitiques et schisteuses hébergeant les minéralisations. Compte tenu du fait que la plupart des graviers minéralisés et des indices minéraux ont été identifiés au niveau des zones de contact schistes – granites définies par Cruys, (1965) et notamment, entre les formations de granites à biotites subordonnées et celles des schistes sériciteux et chloriteux ou des schistes argileux (Allou, 2005) la répartition spatiale des nouvelles zones potentiellement prospectives, sera analysée en fonction de cette superficie d'intérêt géologique. Enfin, les accidents majeurs (failles, fractures) mentionnés sur la carte sont utilisés pour valider le système de fracturation de la zone d'étude.

3.2.2 Données hydrologiques

Les travaux de prospection réalisés sur la zone d'étude montrent que cette région, du fait d'une géomorphogenèse exosphérique assez prononcée, est actuellement dans sa presque totalité, recouverte par des dépôts éluvionnaires, colluvionnaires et alluvionnaires qui occupent la quasi-totalité de la surface. La littérature mentionne également que, les placers hébergeant les minéralisations connues de Nb-Ta de la région, se sont formés à la suite de plusieurs processus hydromorphologiques, marqués par d'importantes phases de drainage et d'alluvionnement subséquentes à une phase de fragmentation et d'usure des roches filoniennes d'altération hydrothermale de natures diverses (Cruys, 1965 ; Adam, 1968).

Partant de ce constat, des données hydrologiques sont utilisées pour extraire des paramètres susceptibles d'influencer le transport mécanique des sédiments et des minéraux lourds au sein des sous-bassins versants étudiés. Ce qui permet de caractériser des zones favorables ou défavorables à l'accumulation de graviers minéralisés. Ces paramètres permettront par la suite de cartographier la capacité des eaux de ruissellement à mobiliser, transporter,

éroder et déposer des graviers minéralisés sur une surface donnée. Ces paramètres permettront également d'identifier des zones favorables à la formation de pièges naturels.

Les données hydrologiques proviennent d'un modèle numérique d'élévation (MNE) ASTER (ASTGTM2_N06W007_dem) acquis à 30 m de résolution le 17 octobre 2011 et couvrant la zone d'étude. L'image est présentée à la figure 12.

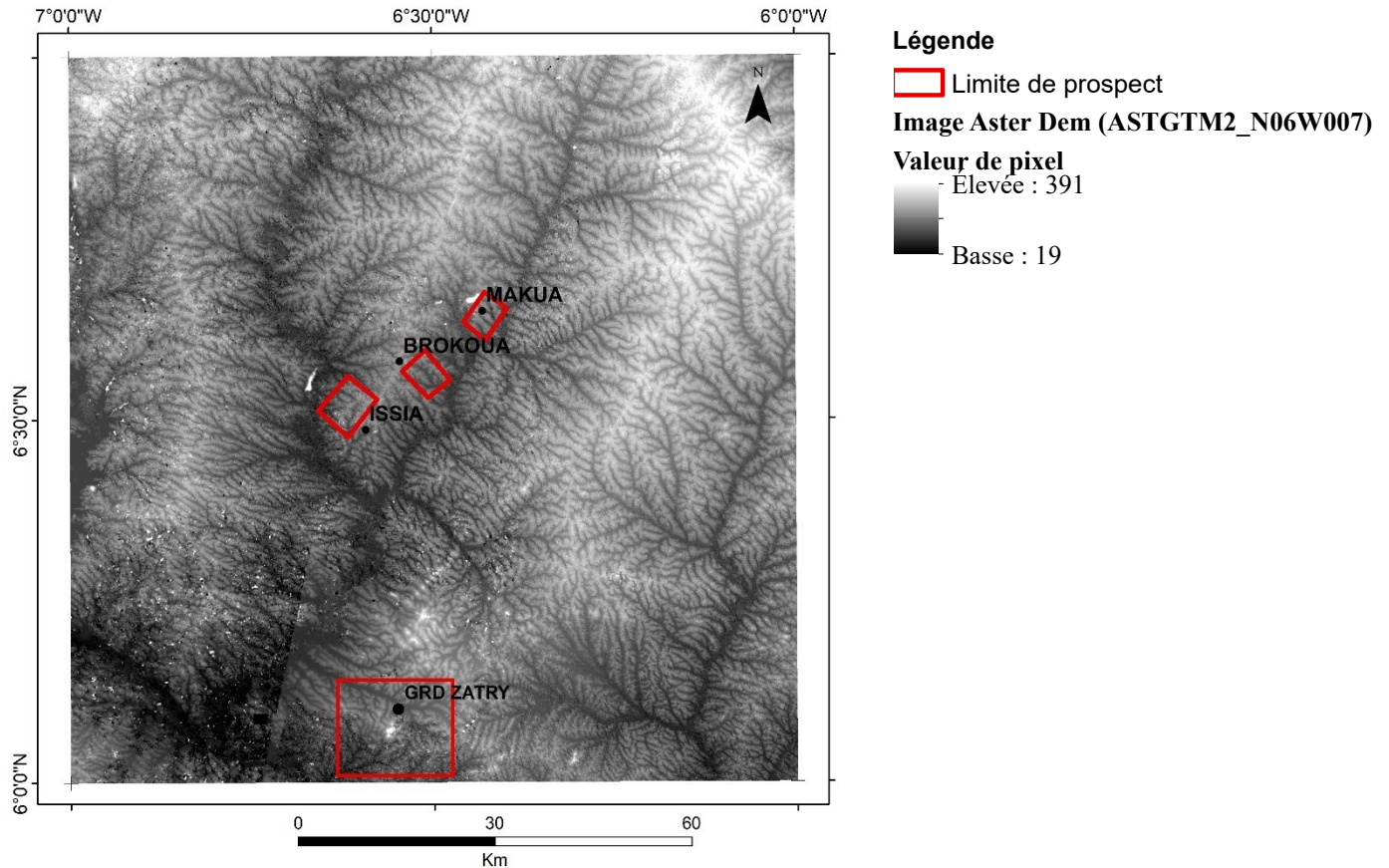


Figure 12: Image satellite Aster DEM (ASTGTM2_N06W007)

L'information contenue dans les pixels est utilisée pour extraire les paramètres de surfaces d'accumulation et d'amplitude de flux ainsi que le réseau hydrographique. En effet, les surfaces de captation de l'eau de ruissellement et le réseau à travers lequel l'eau circule jusqu'à un orifice sont appelés "système de drainage". Le flux d'eau à travers le système de drainage représente un sous-ensemble du "cycle hydrologique". Les surfaces d'amplitude et d'accumulation de flux représentent la variation spatiale du phénomène d'écoulement des eaux, paramétré suivant une valeur au niveau de chaque pixel du modèle numérique d'élévation, couvrant l'étendue de la zone d'étude. Ces paramètres sont susceptibles

d'influencer les propriétés mécaniques et chimiques des graviers minéralisés et de contribuer dans le temps et dans l'espace à l'accumulation de sédiments minéralisés au niveau des zones de pièges naturels.

3.2.3 Données morphologiques

Les travaux de prospection minière réalisés en 2015, ont mis en évidence plusieurs zones favorables à la concentration de minéraux lourds de coltan. Ces zones sont localisées soit au niveau des sommets et des flancs de pénéplaines, soit à proximité d'affleurements rocheux de pegmatites granitiques minéralisées peu altérées, ou soit dans les vallées et les zones de flats. Ce qui signifie que la forme du relief peut exercer une influence sur la répartition spatiale des dépôts connus.

Une première catégorie de données morphologiques est utilisée pour représenter et cartographier le relief local. Ce qui permet d'étudier la répartition topographique des différents dépôts connus de la région étudiée. Le paramètre morphologique sélectionné pour représenter et cartographier la variation spatiale du relief est l'altitude exprimée en mètre.

Une deuxième catégorie de données morphologiques est utilisée pour représenter et cartographier les facteurs morphologiques susceptibles d'influencer ou de contrôler les processus de drainage et de dépôts de minéraux lourds au niveau des sous-bassins versants étudiés. Il s'agit des surfaces de courbure des pentes et du degré d'inclinaison des pentes. En effet, une courbure de pente longitudinale affecte l'accélération et la décélération des eaux de ruissellement, tandis qu'une courbure « planiforme » en influence la convergence et la divergence. L'utilisation de ces paramètres comme indicateurs spatiaux de favorabilité ou de défavorabilité au drainage, à l'érosion ou la sédimentation des minéraux lourds, permet de cartographier et d'identifier des zones d'irrégularités topographiques. Lesquelles sont susceptibles d'influencer la concentration ou le transport mécanique de minéraux lourds et de générer des gîtes exploitables au sein des zones de pièges naturels. Tous ces paramètres sont extraits du même modèle numérique d'élévation (MNE) ASTER (ASTGTM2_N06W007_dem) couvrant la zone d'étude (figure 12).

3.2.4 Données structurales

Les fractures et les fentes de traction générées par la déformation transcurrente (D2), contribuent à la migration des fluides minéralisateurs responsables de la mise en place de pegmatites granitiques minéralisés. À l'échelle régionale, les failles principales sont orientées NS, NE, NW et EW. Les données structurales (failles, fractures) sont utilisées pour cartographier les zones de faiblesses du socle et étudier l'influence du système de fracturation régional sur la localisation des roches pegmatitiques minéralisées primaires d'une part, et sur la répartition des dépôts secondaires (produits d'altération des pegmatites minéralisées) d'autre part. Les données structurales sont extraites à partir de plusieurs images satellites. Il s'agit d'une image multispectrale *Sentinel 2B*- MSI de Niveau 1C (acquises le 24 décembre 2017 en orbite descendante) présentée à la figure 13.

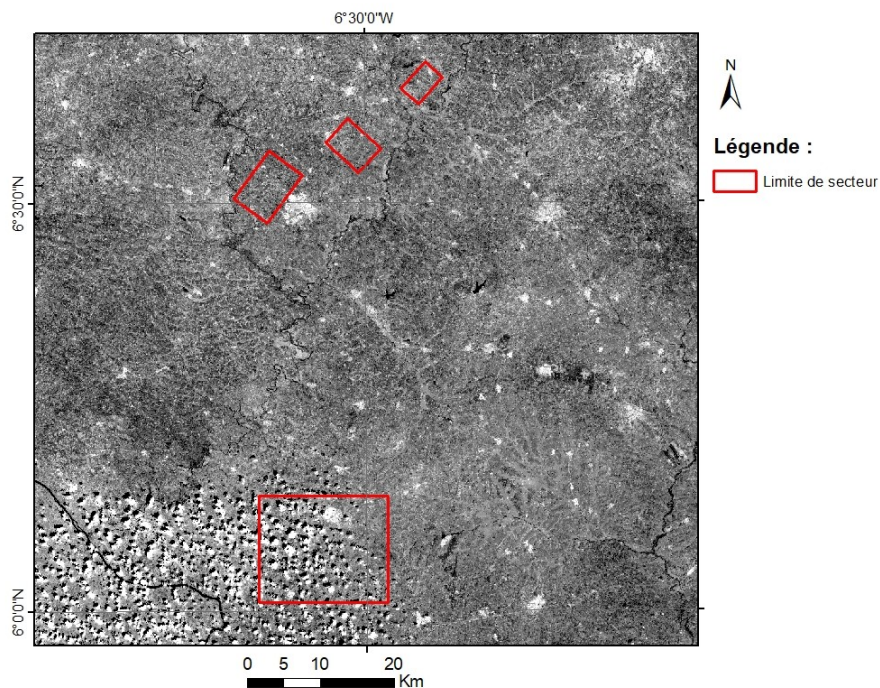


Figure 13: Mosaïque d'images multispectrales *Sentinel-2B* MSI L1C

L'image multi spectrale couvre 13 bandes de longueurs d'onde (443 à 2190 nm), avec une largeur de bande de 10 m pour les bandes B02 (Bleue), B03 (Verte) et B04 (Rouge) du visible et du proche infrarouge B08 (NIR); 20 m pour les bandes B05, B06, B07 et B08A

du très proche infrarouge et de l'hyper spectrale B11 (SWIR-1) et B12 (SWIR-2). Les bandes de correction atmosphérique B01, B09 et B10 sont à 60 m de résolution.

L'image radar *Sentinel-1A* (acquises le 6 janvier 2018 en orbite descendante en mode interférométrique large) est composée de bandes d'amplitudes et d'intensités de polarisation simple HH et VV de résolution 40 m pour 22 x 5 regards et est présentée à la figure 14.

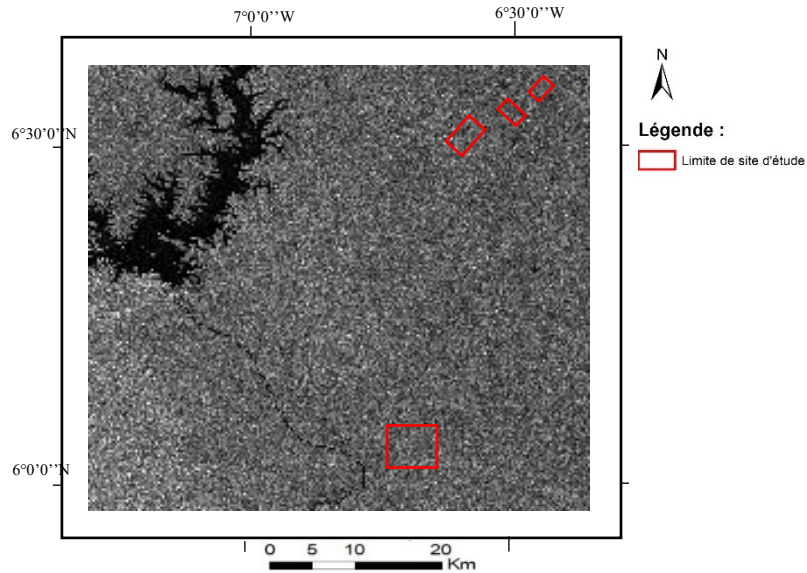


Figure 14 : Image radar Sentinel 1A SAR IW GRD

Ces images ont été utilisées pour extraire des linéaments naturels et cartographier le système de fracturation de la zone d'étude. Une mosaïque d'imagerie de haute résolution spatiale et libre d'accès fournit des images satellites et aériennes d'un mètre ou plus dans de nombreuses régions du monde.

La mosaïque couvrant la zone d'étude a été téléchargée depuis la base de données ESRI est présentée à la figure 15.

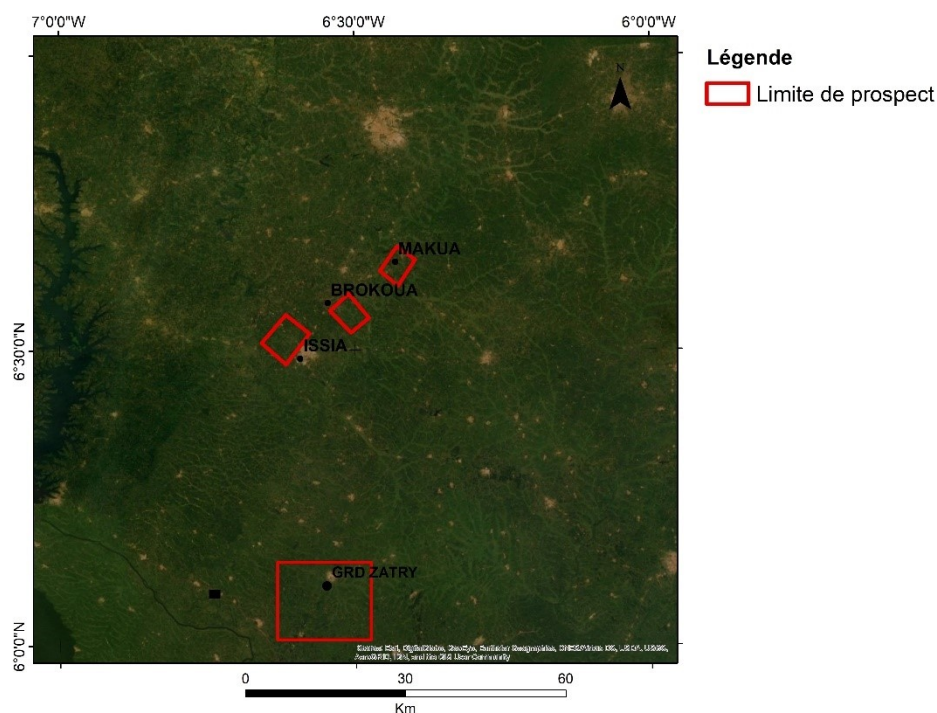


Figure 15 : Imagerie haute résolution Terra Color

Elle se compose d'images Terra Color de quinze mètres à petite et moyenne échelle (591 à 72 m) et des images SPOT de 2,5 m (288 à 72 km). Ce système a été utilisé pour étudier l'apparence à haute résolution des objets dans la zone d'étude, dans le cadre de la validation des linéaments naturels.

3.2.5 Données de prospection minière

Les données de prospection minière utilisées dans cette étude proviennent de 3709 échantillons de sols prélevés au sein des zones de contact schistes – granites, au moyen de puits de dimension 1,20 m de long par 0,80 m de large. La profondeur maximale de ces puits n'excède pas trois (03) mètres. À l'exception des puits de reconnaissance, réalisés lors de notre collecte de données dans la partie centre du Complexe, tous les autres puits utilisés dans cette étude (notamment dans la partie nord du Complexe) ont été positionnés suivant une grille à maille régulière de 100 m de large par 50 m de long. Pour chaque puits, deux couches de graviers de granulométrie différentes ont été échantillonnées au pan. Les échantillons de graviers ont été lavés au « *sluice* », puis la masse de concentré en minéraux lourds (coltan) a été extraite à la batée. Le paramètre minéralogique utilisé pour étudier la

distribution spatiale des occurrences minérales est la teneur moyenne de coltan calculée pour deux échantillons par puits et exprimée en gramme par mètre cube. Les teneurs varient de 0 à 1157 g/m³. Le jeu de données de prospection minérale utilisé est composé de trois (03) champs. Les deux premiers champs sont les coordonnées géographiques (*easting*, *northing*) de l'emplacement des puits et le troisième est la valeur de teneur moyenne en coltan échantillonnée dans les puits.

3.2.6 Données administratives et géominières

Les données du découpage administratif de la Côte d'Ivoire et les données géominières (géologie, géophysiques, minéralogie) sont issues des bases de données à référence spatiale du Ministère de la Géologie et des Mines de Côte d'Ivoire et du Bureau national d'étude et de travaux de développement (BNTED). Ces données sont utilisées comme couches thématiques dans l'application cartographique SIG *Web* et permettront d'informer les utilisateurs de l'outil sur la disponibilité de l'information géominière, utilisable pour guider des travaux de prospection notamment sur l'accessibilité des zones prospectives potentielles. L'affichage d'une base de données documentaires (rapports d'exploration) relative aux indices miniers peut être utilisée pour contribuer à améliorer les connaissances acquises sur les systèmes de minéralisation. L'affichage des données géologiques dans l'outil renseignera le géologue sur les grands ensembles géologiques hébergeant les zones à prospecter. L'affichage des données géophysiques aéroportées (magnétiques et radiométriques) renseigneront sur la disponibilité des cartes de levés géophysiques aux échelles 1/50 000^{ème} et au 1/200 000^{ème} (lorsqu'elles sont disponibles). Ces données peuvent être acquises et traitées par le géologue en vue d'améliorer ses connaissances sur les propriétés magnétiques ou radiométriques des métaux. Pour des raisons de confidentialité et d'accès aux données, ces types de cartes géophysiques n'ont pu être présentés dans cette étude. Des cartes de fonds (*Open street map* et *Bing map*) sont intégrées dans l'outil pour renseigner sur la topographie, le réseau de transport et l'occupation des sols. L'affichage des données administratives dans l'outil peut renseigner le prospecteur sur les subdivisions territoriales et l'orienter vers les directions régionales ou départementales (du ministère des Mines et de la Géologie) administrant les superficies potentiellement prospectives dans le cas de la Côte d'Ivoire.

MÉTHODE DE TRAITEMENT DES DONNÉES

Compte tenu de l'insuffisance des données nécessaires pour étudier les relations spatiales, existantes entre des indicateurs de minéralisations et les indices minéraux de coltan connus; l'approche choisie utilise la puissance prédictive des algorithmes de classification automatique pour modéliser et cartographier la prospectivité de nouvelles cibles minérales. L'ensemble de la démarche méthodologique suivie s'articule sur huit étapes.

La première étape consiste à réaliser une analyse statistique d'autocorrélation ou d'agrégation spatiale des teneurs moyennes de puits, extraites des données de prospection minérale. Cette étape, permet d'étudier non seulement, le mode de distribution géostatistique des graviers minéralisés échantillonnés, mais aussi de discriminer et d'identifier pour la zone d'étude, différents types de cibles minérales.

La deuxième étape consiste à prétraiter dans un environnement SIG les données hydromorphologiques et structurales. Ce qui permet d'extraire et de cartographier les facteurs sous-jacents explicatifs susceptibles de contrôler et d'influencer les processus de drainage et de dépôt de sédiments, au niveau des sous-bassins versants. Cette étape permet également de cartographier et d'étudier la distribution spatiale des pegmatites minéralisés en certains endroits du relief.

La troisième étape consiste à extraire et à attribuer les valeurs des paramètres hydromorphologiques et structuraux aux différents emplacements occupés par chaque entité de puits. Par la suite, l'ensemble de ces paramètres est utilisé comme critère de preuve et comme prédicteurs spatiaux dans l'étape de modélisation prédictive. Ces prédicteurs spatiaux sont utilisés pour exprimer la favorabilité de certaines zones au drainage, à l'érosion et à la formation de différents types de cibles minérales.

La quatrième étape consiste à construire une table de jeu de données d'apprentissage (avec 80 % des données) et réserver les 20 % restant pour la validation des modèles prédicteurs.

La cinquième étape est celle de la modélisation prédictive proprement dite. Cette étape est réalisée en utilisant plusieurs algorithmes de classification automatique (KNN, SVM et DT) et permet de construire plusieurs modèles prédicteurs.

La sixième étape permet, dans un premier temps d'évaluer les performances d'apprentissage des algorithmes en termes de précision de classification globale, de matrice de confusion, de valeurs d'aire sous la courbe, de F-score et de coefficient kappa. Cette étape permet d'identifier les trois meilleurs modèles prédicteurs de chaque série d'algorithmes entraînés (KNN, SVM et DT). La validation de l'apprentissage permet d'évaluer avec un jeu de donnée externe à leurs entraînements, la robustesse de chaque meilleur modèle prédicteur.

La septième étape consiste à sélectionner et utiliser le modèle le plus performant et le plus robuste des trois catégories, pour cartographier dans un environnement SIG, de nouvelles zones prospectives potentielles.

Pour finir la huitième étape a consisté à réaliser un géoportail suivant une architecture SIG *Web Open source*. Les cartes résultantes de prospectivité potentielle du coltan ont alors été intégrées dans l'outil d'aide et l'information géominière est diffusée de façon démocratique et au plus près de la source. L'ensemble de cette démarche méthodologique peut être résumée dans l'organigramme suivant présenté à la figure 16.

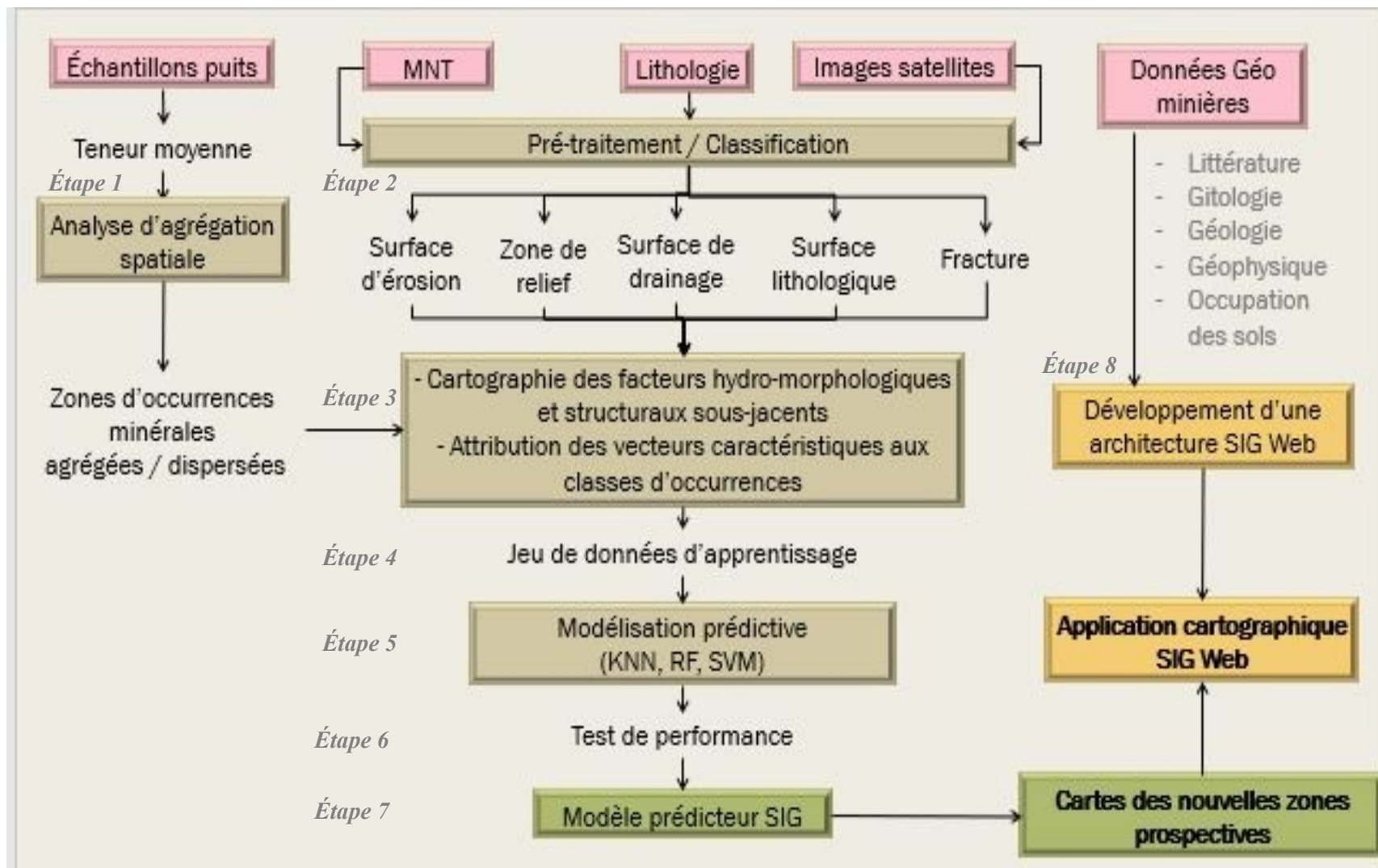


Figure 16 : Organigramme méthodologique général.

3.3 Méthode de discrimination des cibles minérales

La représentation cartographique d'une série statistique est une étape très importante dans l'analyse d'un problème géospatiale. Car, elle donne une information sur la forme, la nature et le mode de distribution spatiale de la population d'échantillon observée. Les objets étudiés dans ce projet de recherche sont les graviers minéralisés en coltan. Le paramètre utilisé pour caractériser la distribution géostatistique de ces graviers minéralisés sur les sites d'étude est la teneur moyenne en concentré de coltan échantillonnée dans chaque puits.

La technique d'analyse d'autocorrélation spatiale dans un SIG, aide à déterminer et comprendre le degré de ressemblance qui existe entre un objet étudié et d'autres objets de même nature situés dans son voisinage. Cette analyse du mode de répartition se réalise en fonction d'une superficie donnée. Si cette répartition est aléatoire, l'analyse statistique spatiale s'arrête là, car il n'existerait aucun facteur sous-jacent susceptible de contrôler la répartition spatiale de ces objets étudiés à l'échelle de la population observée. En revanche, si ces objets présentent une répartition anormale, dispersée ou regroupée, de nouveaux tests statistiques doivent être établis pour déterminer les facteurs qui contrôlent cette répartition 2D (Deveaud, 2015).

L'approche a consisté à utiliser la technique d'analyse d'agrégation statistique d'Anselin Moran (Moran, 2010), autrement connue sous le terme, analyse d'autocorrélation spatiale de Moran. Cette approche statistique permet de déterminer le mode d'agrégation (dispersée, aléatoire ou agrégée) d'entités ou de groupe d'entités au sein d'une population. En fait, l'indice I de Moran mesure simultanément l'autocorrélation spatiale des attributs d'entités, en fonction de leurs emplacements et de leurs valeurs respectives. Ce qui permet d'identifier différents types d'agrégats spatiaux présentant des valeurs élevées ou faibles et de discriminer des points spatiaux aberrants (Moran, 2010). Ainsi, une valeur positive de I, indiquera si une entité est voisine d'autres entités présentant des valeurs attributaires similaires élevées ou faibles. Dans ce cas, cette entité fait partie d'une grappe spatiale. On distinguera alors une grappe spatiale de valeurs élevées d'une grappe spatiale de valeurs faibles.

En revanche, une valeur négative de l'indice « I » indiquera si l'entité est voisine d'autres entités présentant des valeurs différentes. Dans ce cas, l'entité est qualifiée de valeur ou de point aberrant(e).

La logique mathématique de cette analyse s'appuie sur le calcul de la moyenne et de la variance de l'attribut de l'entité à évaluer. Il faut noter par ailleurs qu'une différence importante existe entre les techniques d'analyse statistique spatiale et les techniques d'analyse statistique classiques. En effet, les premières nécessitent d'intégrer à l'analyse la notion d'espace. Pour se faire, elles intègrent les relations spatiales directement dans leurs formules mathématiques. En conséquence, un grand nombre d'outils SIG nécessitent (avant l'analyse) de sélectionner une valeur de distance, pour conceptualiser les relations spatiales existantes entre les différentes entités du jeu de données d'entrée. À cet effet, des techniques de conceptualisation spatiale courantes, basées sur le calcul de l'inverse de la distance ou sur l'utilisation d'une distance constante ou de la distance entre les K voisins les plus proches, peuvent être utilisées pour satisfaire à cette condition. Dans cette étude, c'est la technique de l'inverse de la distance qui a été utilisée pour réaliser le modèle conceptuel des relations spatiales. Cette technique est adaptée pour l'analyse d'agrégation spatiale de points. Elle utilise l'impédance ou la fréquentation de chaque entité en fonction de la distance. Toutes les entités ont alors un impact ou une influence sur les autres entités. De sorte que, plus les entités sont éloignées, plus l'impact ou l'influence est réduit(e).

Le logiciel SIG *ArcMap* a été utilisé pour calculer la valeur de l'indice « *Local Morans I* », un score Z, une pseudo valeur de probabilité p et déterminer un code (COType) pour chaque entité du jeu de données d'entrée. Le champ (COType) fait la distinction entre les types de grappes de points identifiés. Le résultat est soit une grappe statistiquement significative de valeurs élevées (HH ou *High-High*) ou une grappe statistiquement significative de valeurs faibles (LL ou *Low-Low*), soit une grappe de points aberrants (statistiquement non significative) dans laquelle, une valeur élevée est entourée principalement de valeurs faibles (HL ou *High-Low*) ou une valeur faible est entourée principalement de valeurs élevées (LH ou *Low-High*). Les scores Z très élevés (de valeurs positives) ou très faibles (de valeurs négatives) sont associés à des valeurs p très basses. Ces scores Z se trouvent aux extrémités d'une courbe de répartition normale et indiquent

qu'il est peu probable que le modèle spatial observé reflète un modèle aléatoire théorique. Une telle répartition indique que les attributs analysés présentent une distribution structurée au sein de la superficie étudiée et que des processus spatiaux sous-jacents sont à l'œuvre et sont susceptibles d'influencer leur mode de distribution spatiale.

Aussi, il faut noter que la méthode d'analyse d'autocorrélation spatiale de Moran est une technique d'analyse statistique inférentielle. Les résultats de l'analyse doivent être interprétés dans un contexte d'hypothèse nulle. Par exemple, lorsque les pseudo valeurs p des entités composant les grappes de points (à valeurs élevées, faibles ou aberrantes) sont assez basses et inférieures à 0,05, cela signifie que ces points sont considérés comme agrégés statistiquement à une probabilité de 95 %. De ce fait, les entités de notre jeu de données de prospection minérale rejetant l'hypothèse nulle, à savoir les entités ne présentant pas un mode d'agrégation spatiale statistiquement significatif, ont été renommées (AL) (Aléatoire). Les grappes spatiales statistiquement significatives (LL, LH, HH et HL) ont été utilisées pour représenter (dans l'étude) les différents types de cibles minérales (dépôts connus) « prospectables » dans le cadre d'opérations de recherche minière. Les entités non significatives statistiquement (AL) ont été utilisées et retenues dans l'étude pour représenter les zones peu prospectives pour les cibles minérales de type HH, HL, LL et LH.

Une autre condition d'application de cette approche statistique est la présence d'au moins 30 observations ou entités dans le jeu de données d'entrée. Pour cette raison, l'analyse d'autocorrélation spatiale a été appliquée sur les sites de Bémadi, Brokoua et Makua. Mais n'a pu l'être sur les sites de Grand Zatry et Oureyo, compte tenu de l'insuffisance des données de puits collectées lors de notre campagne de terrain. Par conséquent, les résultats cartographiques de cette analyse statistique sont trois cartes présentant la distribution spatiale des cibles minérales connues de la région nord du Complexe géologique Hana-Lobo.

La prochaine étape présente les méthodes utilisées pour extraire les critères de preuves hydromorphologiques et structurales. Il s'est agi d'extraire et de cartographier (à partir de données géospatiales) les paramètres sous-jacents susceptibles d'influencer la distribution spatiale des différents types de cibles minérales discriminées.

3.4 Techniques d'extraction des critères spatiaux de preuves

L'analyse des relations spatiales existantes entre des critères de preuve spatiaux et des emplacements d'occurrences connus peut être réalisée à l'aide de différentes méthodes numériques (Carranza *et al.*, 2008). Il s'est agi ici de prétraiter dans un environnement SIG, les données hydromorphologiques et structurales, en vue d'extraire et de cartographier la variation spatiale de ces paramètres sur toute l'étendue de la zone d'étude. Les surfaces cartographiées seront alors utilisées comme critères spatiaux de preuves sous-jacents, pour caractériser la susceptibilité de ces paramètres à influencer et/ou à contrôler la distribution spatiale des cibles minérales et la formation de dépôts de minéraux lourds au sein des sous-bassins versants de la Lobo.

Dans cette étude, l'échelle de la plupart des cartes de preuves utilisées est au 1/30 000^{ème}. Cette échelle permet de fournir suffisamment de données originales pour obtenir des résultats d'interpolation fiables, qui satisfont à une précision de présentation de caractéristiques probantes, dans le cadre d'une CPM à l'échelle régionale (Sun *et al.*, 2019).

3.4.1 Techniques d'extraction des paramètres hydrologiques

Les paramètres de surface d'accumulation et d'amplitude de flux de drainage, ainsi que le réseau hydrographique, ont été extraits à partir des données géospatiales tridimensionnelles (latitude, longitude, altitude) de chaque pixel du MNE. La cartographie des axes de concentration naturelle des eaux de ruissellement de surface, met en évidence la capacité des eaux de ruissellement à mobiliser, transporter, éroder et déposer des minéraux lourds sur une surface de 30 m x 30 m. Des algorithmes dédiés ont été utilisés dans un SIG, pour calculer les valeurs des surfaces d'accumulation et d'amplitude de flux sur toute l'étendue de la zone d'étude. Ces valeurs sont dérivées des mesures d'élévation pour une surface d'altitude donnée. Les cellules en sortie présentant une accumulation de flux élevés sont des zones de flux concentrés, qui peuvent être utilisées pour identifier des canaux d'écoulement. Les cellules en sortie, présentant une accumulation de flux, faibles à nuls, sont des zones topographiques locales élevées, qui peuvent être utilisées pour identifier des zones de crêtes, de sommets ou de flancs dans les interfluves. Les valeurs des surfaces

calculées ont été extraites aux coordonnées géographiques de chaque entité composant les différents types de cibles minérales.

Ces valeurs ont été assignées dans la table de données de prospection minérale et seront utilisées comme variables prédictives hydrologiques dans le jeu de données d'apprentissage.

3.4.2 Techniques d'extraction des paramètres morphologiques

Les paramètres morphologiques cartographiés sont les surfaces de courbure ou profil de pente, le degré d'inclinaison de pente et la forme du relief. Ces paramètres sont susceptibles d'influencer l'accélération, la décélération, la convergence ou la divergence des eaux de ruissellement à l'échelle des sous-bassins versants étudiés. Les valeurs sont dérivées des données géospatiales tridimensionnelles (latitude, longitude, altitude) de chaque pixel du MNE.

Dans le logiciel *ArcMap*, un algorithme dédié au calcul du taux de variation maximal des valeurs de ces paramètres dans chaque cellule par rapport à ses plus proches voisines a été utilisé. D'une façon générale, la variation maximale de l'altitude, sur une distance séparant une cellule de ses huit voisines, identifie la descente la plus raide depuis cette cellule. Ce qui permet d'identifier le dégradé ou le taux de variation maximal d'altitude (la pente) dans chaque cellule de la surface étudiée et de paramétrer la raideur du terrain. Une surface plane est à 0 %, une surface à 45 degrés est à 100 %. En somme, plus la surface devient verticale et plus le pourcentage ou le degré d'inclinaison de la pente devient de plus en plus grand.

D'autre part, les valeurs de courbure (concavité ou convexité) des pentes peuvent être calculées cellule par cellule dans un SIG. La valeur de courbure de chaque cellule est alors calculée sur une surface composée d'une fenêtre de 3 cellules sur 3. La sortie générée par l'outil permet ainsi d'identifier les caractéristiques physiques susceptibles d'affecter les processus d'érosion et d'écoulement à l'échelle d'un bassin de drainage.

En outre, la valeur de courbure a permis de rechercher les motifs d'érosion du sol ainsi que la distribution de l'eau sur les surfaces étudiées. Les cartes de surfaces de preuves morphologiques ont été générées sur toute la zone d'étude et les valeurs caractéristiques

calculées ont été extraites aux coordonnées géographiques de chaque entité correspondant aux emplacements des différents types de cibles minérales. Ces valeurs ont été assignées dans la table de données de prospection minérale et seront utilisées comme variables prédictives morphologiques dans le jeu de donnée d'apprentissage.

3.4.3 Techniques d'extraction des paramètres structuraux

Le critère de preuve structural considéré pour cette étude est le système de densité de fracturation. Ce paramètre est susceptible de favoriser la formation d'anfractuosités topographiques ainsi que la mise en place de roches hôtes potentiellement minéralisées (pegmatites granitiques).

Selon Li, (2010), les travaux de Walsh *et al.* (1986), comparant cinq techniques d'amélioration différentes (valeur moyenne de toutes les bandes, Analyse en composantes principales (ACP), taux de bande, égalisation d'histogramme et filtre passe-haut) ont montré que l'ACP est une technique efficace pour l'identification des linéaments. D'autre part, l'étude de Paganelli *et al.* (2003) a permis d'évaluer la pertinence de l'utilisation de la technique d'analyse des composantes principales (ACP) des images radars, pour la détection de caractéristiques linéaires géologiques. Les résultats ont montré que l'utilisation de l'ACP peut fournir une base d'image améliorée pour la cartographie structurale. En effet, les résultats ont montré entre autres que la bande PC2 de l'image RADARSAT-1, optimise la perception topographique inhérente aux images radar et conservent l'information relative à la réponse radar à la surface. Cela améliore notamment, l'interprétabilité des structures linéaires géologiques. Plus tard, les travaux de Adiri *et al.* (2017) ont montré qu'en plus des images *Sentinel* 1, des linéaments géologiques peuvent être extraits des bandes VNIR et SWIR des images OLI et ASTER. La technique utilisée dans notre étude s'appuie sur les travaux de Adiri *et al.* (2017). Elle a consisté à extraire des linéaments à partir des images radar et optiques *Sentinel*. L'ensemble du processus d'extraction automatique et de cartographie de la densité de linéaments peut être résumé dans l'organigramme de la figure 17.

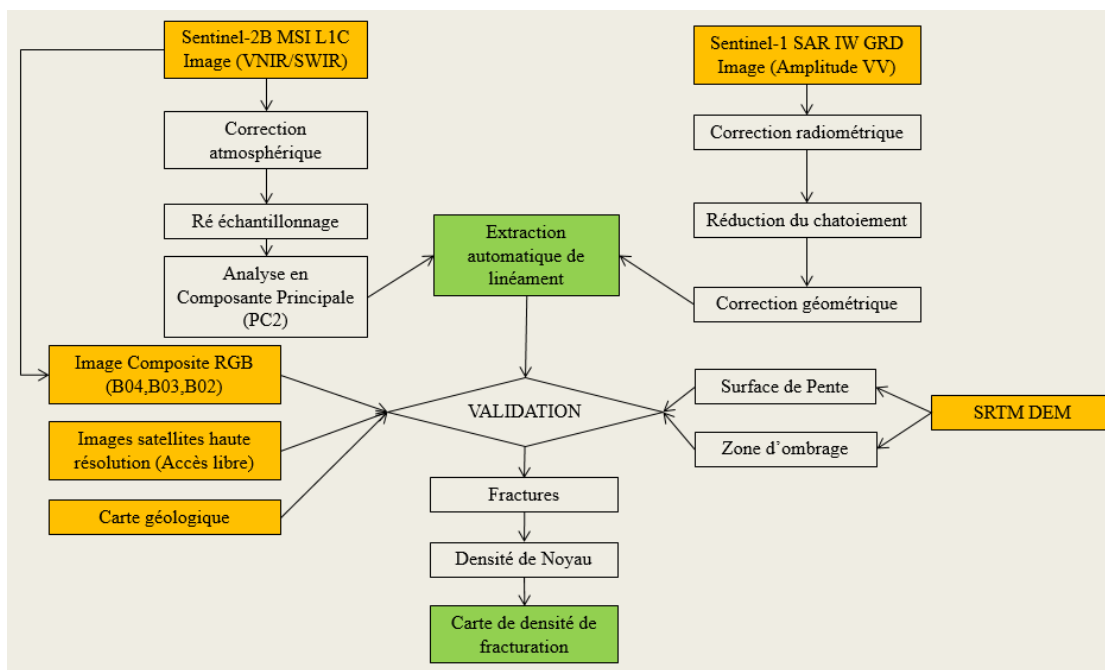


Figure 17 : Organigramme méthodologique d'extraction automatique de linéaments.

Cette nouvelle approche méthodologique a permis de dépasser les procédures classiques habituelles et d'obtenir plus de linéaments pour la zone étudiée. Concrètement, il s'est agi d'extraire de façon automatique des linéaments issus d'images optiques multispectrales (VNIR et SWIR) et radars prétraités puis de valider les linéaments naturels extraits à partir de données de vérité terrain. A cet effet, les bandes B11 (SWIR) et B08 (NIR) ont été rééchantillonnées à 10 m de résolution. La bande C de polarisation VV de l'image *Sentinel-1* SAR IW GRD a été corrigée radiométriquement et géométriquement à 10 m de résolution. Ces bandes ont été utilisées pour extraire de façon automatique des linéaments sur toute la zone d'étude.

Concrètement, le logiciel SNAP a été utilisé pour réaliser une correction automatique des effets atmosphériques contenus dans les scènes *Sentinel-2B* MSI L1C. Ce qui a permis d'obtenir pour les bandes B02, B03, B04 et B08 à 10 m de résolution, des ortho-images de surface de réflectance de niveau 2A. Puis, la bande B11 (rééchantillonnée à 10 m) et la bande B08 ont fait l'objet d'une analyse en composantes principales, de sorte à compresser les informations contenues dans les bandes initiales dans de nouveaux canaux appelés Composants Principaux (PC) (Zhang *et al.*, 2007; Gabr *et al.*, 2010; Adiri *et al.*, 2015).

Cela permet d'éliminer la redondance des données, d'isoler le bruit et d'améliorer ensuite les informations ciblées dans l'image (Amer *et al.*, 2012). Ensuite, une mosaïque d'images radar *Sentinel-1* SAR IW GRD a été construite. Il s'est agi de la calibrer et de la corriger topographiquement en utilisant un modèle numérique STRM 3 arcsec (90 m de résolution) valable pour la zone d'étude. Les valeurs numériques ont été converties en amplitude Gamma0 de polarisation simple VV et le bruit ou chatoiement dans les images a été réduit à l'aide du filtre de Lee dans une fenêtre glissante de 3 x 3. Les images résultantes ont été converties et rééchantillonnées à une résolution de 10 m.

La technique d'extraction automatique de linéaments est basée sur un algorithme de détection automatique de lignes, de contours, d'angles et de distances dans les pixels de l'image. La boîte à outils « *Lineament extraction* » du logiciel PCI *Geomatica* fournit l'algorithme. Les valeurs de ces caractéristiques sont calculées selon plusieurs paramètres.

Les paramètres RADI (*filter radius*) et GTHR (*Edge Gradient Threshold*) sont utilisés pour la détection de contours. Les paramètres LTHR (*Curve Length Threshold*) et FTHR (*Line Fitting Threshold*) pour la détection des lignes et les paramètres ATHR (*Angular Difference Threshold*) et DTHR (*Distance Linking Threshold*) pour la détection des angles dans les pixels de l'image. Les travaux menés par Corgne *et al.* (2010) et Hashim *et al.* (2013) recommandent d'utiliser une longueur minimale de 10 pixels comme seuil de détection LTHR acceptable ; une valeur comprise entre 3 et 8 pixels comme rayon de filtre RADI pour éviter l'introduction de bruit dans le processus ; une valeur comprise entre 10 et 70 pixels comme niveau minimum de détection de bordure GTHR et une valeur comprise entre 2 et 5 comme seuil maximal de tolérance FTHR pour ajuster une courbe en polyligne. Également, une valeur comprise entre 3 et 20 degrés est recommandée comme angle maximal ATHR à ne pas dépasser pour une liaison de deux polygones. Enfin, une valeur comprise entre 10 et 45 pixels comme distance maximale DTHR est acceptable pour lier deux polygones. Par conséquent, les paramètres suivants ont été utilisés dans le cadre de cette étude : RADI : 5; GTHR : 30; LTHR : 5; FTHR : 5; ATHR : 30 et DTHR : 30. Le résultat est une carte des linéaments naturels et androgènes couvrant toute la zone d'étude.

Les linéaments naturels ont par la suite été discriminés, puis validés à l'aide de la littérature et des images satellites haute résolution disponibles sur la zone d'étude. Les cartes de

densité de linéaments validés et une rosace directionnelle régionale ont été construites dans un SIG.

Les valeurs de densités obtenues ont été extraites aux coordonnées géographiques de chaque entité composant les différents types de cibles minérales. Ces valeurs ont été assignées dans la table de données de prospection minérale et seront utilisées comme variables prédictives structurales dans le jeu de donnée d'apprentissage.

3.4.4 Techniques de validation des linéaments naturels

Dans l'optique de discriminer les linéaments naturels dans cet ensemble, une étape de validation s'est avérée nécessaire. En effet, les travaux d'Hashim *et al.* (2013) ont montré que la combinaison d'opérations de traitement d'images numériques, basées sur les contours et les liaisons de lignes avec des paramètres optimaux locaux, est a priori cruciale pour extraire des caractéristiques de linéaments dans les régions à forte végétation. Étant donné que notre zone d'étude se situe dans un contexte de climat tropical humide à forte végétation, les données de télédétection acquises par satellite peuvent être vulnérables au niveau de bruits générés dans la dimension des pixels. Ces niveaux de bruits sont la résultante de l'effet de diffusion par les nuages ou de la proximité des couvertures végétales fortement hétérogènes. Dans ce cas, des informations auxiliaires (image de très haute résolution spatiale, carte topographique ou géologique détaillée, données de terrain, etc.) peuvent être combinées, pour distinguer des linéaments androgènes extraits des linéaments naturels de façon automatique (Saadi *et al.*, 2011; Nkono *et al.*, 2013; Hashim *et al.*, 2013).

L'étape de validation des linéaments naturels a donc nécessité de croiser plusieurs informations fournies par la littérature avec celles issues de l'observation d'images de haute résolution spatiale. À cet effet, une carte d'aspect d'illumination (*shading*) des pixels voisins a été générée à partir du MNE. L'illumination dans les pixels, les accidents tectoniques majeurs issus de la carte géologique régionale au 1/200 000 et l'occupation des sols, fournie par un système d'imagerie haute résolution libre d'accès, ont été conjointement utilisés pour discriminer les linéaments extraits en deux catégories (naturels et androgènes). Les linéaments naturels sont les lignes de fractures naturelles, le réseau hydrographique ou les contours géologiques. Les linéaments androgènes sont les limites

de terres cultivées ou de forêts, les limites de zones bâties et les voies de transports terrestres ou ferroviaires et souvent les lignes électriques de haute tension visibles sur des images de haute résolution spatiale.

Une fois, les linéaments naturels extraits, plusieurs méthodes d'analyse statistique de linéaments sur une zone géographique peuvent être utilisées afin d'étudier la géométrie du réseau de linéaments et d'en dégager les directions dominantes à l'échelle régionale (Pretorius *et al.*, 1974; Deslandes *et al.*, 1991 in Coulibaly, 1996; p.30). La méthode conventionnelle consiste à produire des rosaces directionnelles proportionnelles à la longueur cumulative des linéaments par classes de 10 degrés d'orientation. L'analyse statistique des directions majeures et préférentielles permet de mettre en évidence le système de fracturation de la zone étudiée. Par conséquent, une rosace directionnelle a été construite à l'aide d'un progiciel spécialisé (*RockWorks*) et les directions régionales préférentielles ont été analysées, puis validées au regard de l'information et des observations acquises sur la zone d'étude.

3.4.5 Technique de cartographie de la densité de fracturation

En théorie, une surface courbe et lisse passe par au moins une ligne. Sa valeur est la plus élevée sur la ligne et diminue à mesure que l'on s'en éloigne pour atteindre une distance de rayon de recherche égale à zéro à partir de la ligne (Silverman, 1986). La densité de fracturation de chaque cellule du *raster* en sortie peut être alors calculée dans un SIG, en ajoutant les valeurs de toutes les surfaces de noyau qui chevauchent le centre de la cellule. Les valeurs de densité ainsi calculées peuvent être cartographiées pour l'ensemble de la zone d'étude.

Ces valeurs ont été extraites aux coordonnées géographiques de chaque entité composant les différents types de cibles minérales. Elles ont été assignées dans la table de données de prospection minérale où elles seront utilisées comme variables prédictives structurales dans le jeu de donnée d'apprentissage.

3.5 Méthode de modélisation prédictive par apprentissage automatique

Cette section du rapport décrit la procédure utilisée pour modéliser de façon prédictive dans un environnement SIG, la distribution géostatistique des cibles minérales connues en fonction des prédicteurs spatiaux. La procédure commence par la construction d'un jeu de données d'apprentissage, se poursuit par la présentation de la technique utilisée pour la modélisation prédictive proprement dite et est suivie de celle utilisée pour les tests de performance (sur les données d'entraînement) et de robustesse (sur les données de validation). La méthodologie générale s'achève avec la présentation de la technique utilisée pour cartographier les nouvelles zones prospectives sur l'ensemble de la zone d'étude.

3.6 Technique de construction du jeu de données d'apprentissage

Une façon d'obtenir une estimation réaliste d'une classification automatique, consiste à mettre de côté une certaine proportion des observations initiales, de trouver des fonctions de classification avec les autres observations, puis d'effectuer le classement des observations mises de côté. Le jeu de données d'apprentissage est généralement composé de 60 à 80 % des observations d'un jeu de données initial (Peters *et al.*, 2007; Toth *et al.*, 2013; Chen *et al.*, 2017). Dans cette étude, le jeu de données d'apprentissage a été construit avec une partition aléatoire de 80 % des données initiales (soient 2967 points) et les 20 % (soient 742 points) ont été réservés pour les tests de validation et de robustesse des modèles prédicteurs en sortie.

Le jeu de données d'apprentissage se compose de six (06) champs de critères spatiaux de preuves et d'un champ cible de grappe spatiale codé (CoType). Le champ CoType est argumenté d'une valeur codée de 1 à 4, qui correspond aux codes des 4 classes de cibles minérales discriminées. Le code 1 est attribué à la classe AL, le 2 à la classe LL, le 3 à la classe LH et le 4 à la classe mixte HH/HL. Les valeurs des paramètres de critères probants extraites, à savoir le degré d'inclinaison de pente, la courbure de pente, la surface d'accumulation de flux, la surface d'amplitude de flux, l'élévation et la densité de fracturation, sont utilisées comme variables prédictives dans le processus de modélisation prédictive. Les valeurs codées du champ de cibles minérales sont utilisées pour traduire et représenter l'appartenance de chaque entité (teneur moyenne de coltan) à un type d'agrégat

spatial (cible minérale). Ces valeurs codées ont été utilisées comme variable de réponse dans les opérations de classifications par apprentissage automatique.

3.7 Technique d'entraînement des classifieurs automatiques

L'approche de modélisation par classification automatique permet de résoudre des problèmes binaires ou multiclassés, de former, valider et ajuster de manière interactive plusieurs modèles basés sur différents types de classifieurs. Cette approche de formation simultanée de plusieurs classifieurs sur un même jeu de données, permet de comparer les performances de chaque modèle prédictif côte à côte et de déterminer le plus précis et le plus performant. La détermination des paramètres à utiliser pour l'entraînement des algorithmes de classification automatique est une étape cruciale dans un processus de modélisation prédictive. Cette étape consiste à élaborer une procédure clé de classification en vue d'obtenir des prédictions fiables. Cependant, il est difficile de spécifier a priori des procédures de configuration appropriées avec la précision souhaitée, car il existe moins de règles pour déterminer des paramètres optimaux dans le monde réel (Sun *et al.*, 2019).

Dépendamment de la structure interne du jeu de données d'apprentissage en entrée, plusieurs scénarii peuvent être envisagés. Une boîte à outils (*LearnerApp*) du logiciel de calcul *Matlab*, a été utilisée à cet effet, dans le cadre de cette étude. L'outil fournit une variété d'algorithmes d'apprentissage automatique non supervisés. Il s'agit notamment, des algorithmes de machines à vecteurs de support (SVM), des arbres de décision (DT) uniques, des algorithmes d'ensembles d'arbres de décision renforcés et ensachés et des algorithmes des k-voisins les plus proches (KNN). La première étape a consisté à entraîner le jeu de données d'apprentissage avec différents modèles construits à partir de plusieurs algorithmes SVM, KNN et DT. Le tableau 2 présente les différentes méthodes, fonctions et paramètres utilisés pour la formation de ces algorithmes. Le résultat d'une classification automatique, réalisée avec un algorithme d'apprentissage automatique, est un modèle prédictif capable d'être utilisé dans une étude de CPM. Les résultats obtenus ont permis de déterminer, pour chaque série d'algorithmes entraînés, lequel était le plus précis et le plus performant pour prédire, de la façon la plus efficace, les quatre types de classes cibles minérales identifiées dans l'étude. De façon pratique, l'application de ces paramètres a permis de jouer sur la flexibilité et l'interprétabilité des différents modèles en sortie.

Tableau 2 : Paramètres des algorithmes d'apprentissage automatique

Catégorie de modèle	Nom du modèle	Paramètres d'apprentissage	Valeur
SVM	Méthodes multi classe :		<i>One-vs-All, One-vs-One</i>
	Échelle de noyau		1
	Niveau de contrainte de boîte :		1
	SVM linéaire		
	SVM à noyau quadratique	Quadratique	
	SVM à noyau cubique	Cubique	
	SVM à noyau gaussien fin	Fonction de noyau : Gaussien, fine	0,61
	SVM à noyau gaussien moyen	Gaussien, moyenne	2,4
KNN	SVM à noyau gaussien grossier	Gaussien, grossière	9,8
	KNN fin	Fine	1
	KNN moyen	Méthode de séparation : Moyenne	10
	KNN grossier	Grossière	100
	KNN cosinus	Méthode de détection et métrique de distance : Moyenne, cosinus	10
	KNN cubique	Moyenne, cubique (Minkowski)	10
	KNN pondéré-égal	Pondérée, égale	1
	KNN pondéré-euclidienne	Pondérée, euclidienne	10
	KNN pondéré inverse distance	Métrique de distance, fonction de pondération : Pondérée, inverse de la distance	100
	KNN pondéré inverse distance carrée	Pondérée, inverse de la distance au carré	100

Tableau 2 : Paramètres des algorithmes d'apprentissage automatique (suite et fin)

Catégorie de modèle	Nom du modèle	Paramètres d'apprentissage		Valeur
Arbre de décision unique (DT)		Fractionnement de la décision de substitution:		Off
	DT simple		Simple, indice de diversité de Gini, règle Twoing, réduction maximale de la déviance	4
	DT moyen	Nature de l'arborescence et critères de fractionnement:	Moyenne, indice de diversité de Gini, règle Twoing, réduction maximale de la déviance	20
	DT complexe		Complexe, indice de diversité de Gini, règle Twoing, réduction maximale de la déviance	100
Ensemble d'arbres de décision	Ensemble DT renforcés	Méthode d'assemblage :	<i>AdaBoost</i> (arbres renforcés)	
			Nombre maximum de divisions :	10-20-30
			Nombre d'apprenants :	30 à 800
			Taux d'apprentissage :	0,1- 1
	Ensemble DT ensachés (RF)		Dimension de sous espace :	1
		Méthode d'assemblage :	<i>Bag</i> (arbres ensachés)	
			Nombre maximum de divisions :	2 à 20
			Nombre d'apprenants :	30 à 800
			Taux d'apprentissage :	0,1- 1
			Dimension de sous espace :	1

3.7.1 Paramètres d'apprentissage des classifieurs SVM

Un modèle SVM linéaire est peu flexible. Il effectue une séparation simple entre deux classes et est par conséquent facile à interpréter. Les modèles SVM à noyau cubique et quadratique sont moyennement flexibles et par conséquent plus difficiles à interpréter. Un modèle SVM à noyau gaussien fin, est très flexible et effectue une distinction fine et détaillée des différentes classes selon un paramètre d'échelle de noyau fixé à $P^2/4$ (où P est le nombre de prédicteurs). Un modèle SVM à noyau gaussien moyen présente une flexibilité moyenne et effectue une distinction de niveau moyen des classes selon un paramètre d'échelle de noyau fixé à P^2 . Un modèle SVM à noyau gaussien grossier est peu flexible et effectue une séparation grossière entre les classes selon un paramètre d'échelle de noyau fixé à $P^2 \times 4$.

Dans cette étude, les paramètres d'échelle de noyau ont été calculés automatiquement dans le logiciel suivant une procédure heuristique. La procédure heuristique utilise le sous-échantillonnage. Par conséquent, pour reproduire les résultats, un nombre de départ aléatoire doit être défini avant de former le classifieur. D'autre part, plusieurs méthodes ont été appliquées pour réduire les problèmes de classification multi-classes à un ensemble de sous-problèmes de classification binaire, avec un apprenant SVM pour chaque sous-problème. La méthode *One-vs-One* a été appliquée pour former un apprenant pour chaque paire de classes. De cette sorte, chaque apprenant apprend à distinguer une classe de l'autre. La méthode *One-vs-All* a été appliquée pour former un apprenant pour chaque classe de sorte que, chaque apprenant puisse également distinguer une classe de toutes les autres.

Le paramètre de contrainte de boîte est une pénalité de marge et constitue une contrainte dure utilisée dans les équations doubles des algorithmes SVM. En somme, ce paramètre permet de conserver les valeurs des multiplicateurs de Lagrange dans une zone ou une région délimitée. Afin d'ajuster un classificateur SVM, le paramètre du niveau de contrainte de boîte peut être augmenté à l'aide d'une valeur scalaire positive. Ainsi, l'augmentation du niveau de contrainte de boîte peut réduire le nombre de vecteurs de support, mais peut cependant, augmenter le temps de l'apprentissage. Tous les apprentissages avec les SVM ont été effectués suivant un mode de données standardisées.

Ce mode est utilisé pour redimensionner chaque distance de coordonnées, dans le cas où les prédicteurs présenteraient des échelles très différentes. Cette normalisation permet en outre d'améliorer l'ajustement des modèles.

3.7.2 Paramètres d'apprentissage des classifieurs KNN

Les classificateurs KNN ont généralement une bonne précision prédictive dans les dimensions faibles, mais pas dans les dimensions élevées. Ils utilisent beaucoup de mémoire et ne sont pas faciles à interpréter. Le paramètre de distance métrique est utilisé pour déterminer et calculer la distance entre les points. Une distance métrique euclidienne utilise la distance séparant deux points, tandis qu'une distance métrique pondérée utilise plusieurs fonctions de pondération. La fonction distance égale, fait fi de toute pondération contrairement aux fonctions inverses de la distance et inverse de la distance au carré.

Le paramètre de voisinage est le nombre de voisins les plus proches à rechercher pour classifier chaque point lors de la prévision. Plusieurs types de classifieurs KNN ont été appliqués dans cette étude. Il s'agit des classifieurs KNN, fin, moyen, grossier, cosinus et pondéré. Un classifieur fin est un algorithme KNN avec un nombre de voisins faible. *A contrario*, un classifieur grossier est un algorithme KNN avec un nombre élevé de voisins. Ainsi, un modèle KNN fin distingue finement les classes avec un paramètre de voisinage fixé à 1. Un modèle KNN moyen distingue moyennement les classes avec un paramètre de voisinage fixé à 10. Un modèle KNN grossier distingue grossièrement les classes avec un paramètre de voisinage fixé à 100. Le modèle KNN cosinus distingue moyennement les classes en utilisant une métrique de distance cosinus et un nombre de voisins fixé à 10. Un modèle KNN cubique distingue également moyennement les classes, mais avec une métrique de distance cubique et un nombre de voisins fixé à 10. Le modèle KNN pondéré distingue aussi les classes moyennement, mais en utilisant un poids de distance entre les 10 plus proches voisins. Dans cette étude, à l'instar de l'entraînement des SVM, toutes les procédures d'apprentissage avec les KNN ont été effectuées avec des données standardisées.

3.7.3 Paramètres d'apprentissage des classifieurs DT

Deux catégories de classifieurs utilisant des arbres de décision ont été testés dans cette étude. Il s'agit des modèles basés sur un seul arbre de décision et des modèles basés sur des arbres de décision assemblés suivant diverses méthodes.

Les arbres de classification sont formés pour prévoir ou prédire les réponses d'un modèle à un problème de classification posé dans le jeu de données d'entraînement. Pour prédire une réponse, les décisions de l'arborescence évoluent depuis un nœud racine (débutant) jusqu'à un nœud feuille contenant la réponse.

Les modèles d'arbres de décision uniques sont faciles à interpréter et sont rapides pour l'ajustement des modèles et la prédiction des réponses. À cet effet, ils utilisent peu de mémoire, mais peuvent avoir une précision prédictive faible.

Le premier paramètre d'apprentissage modifiable est le « Nombre maximum de divisions ». Il permet de spécifier le nombre maximal de divisions ou le nombre de points de branche utilisé pour contrôler la profondeur de l'arborescence. Un arbre fin avec de nombreuses feuilles est généralement très précis sur les données d'entraînement. Toutefois, il est possible que l'arborescence ne présente pas une précision comparable sur un ensemble de tests indépendants. En revanche, un arbre trop feuillu a tendance à surentraîner et sa précision de validation est souvent bien inférieure à celle obtenue lors de son entraînement. En revanche, un arbre grossier n'atteint pas une précision d'entraînement élevée, mais peut s'avérer être plus robuste, en ce sens que sa précision d'entraînement peut être comparable à celle d'un ensemble de tests représentatifs.

Le deuxième paramètre d'apprentissage modifiable est le « Critère de fractionnement ». Il permet de spécifier la mesure du critère de division, afin de décider quand scinder les nœuds. Plusieurs critères de fractionnement ont été utilisés dans cette étude. Il s'agit de l'indice de diversité de Gini, de la règle de Twoing et de la réduction maximale de la déviance. Ces règles tentent d'optimiser des nœuds purs contenant une seule classe. L'indice de diversité de Gini et le critère de déviance mesurent l'impureté du nœud. En revanche, la règle de Twoing utilise une mesure différente pour décider comment diviser le nœud et où maximiser l'expression, afin d'augmenter la pureté du nœud.

Le troisième paramètre modifiable est le paramètre de « Fraction de la décision par substitution ». Il peut être utilisé pour améliorer la précision des prévisions en cas de valeurs manquantes dans les données d'entraînement. Dans ce cas, l'arborescence de classification recherche au plus 10 fractions de substitution sur chaque nœud de branche. Dans cette étude, les arbres uniques simples, moyens et complexes ont été testés sur le jeu de données d'apprentissage sans fractionner la décision par substitution. Les modèles d'arbre unique fin ont été formés avec un nombre maximum de divisions égal à 1, les modèles d'arbre unique moyen avec un nombre égal à 10 et les modèles d'arbre unique complexe, avec un nombre égal à 100.

La deuxième catégorie d'algorithmes utilisés est un classifieur d'arbres de décisions assemblés (classifieurs d'ensembles). Ces algorithmes sont plus complexes et avancés et combinent les résultats de nombreux apprenants faibles ou simples en un modèle d'ensemble de haute qualité. Les qualités de prédiction dépendent du choix de l'algorithme. Dans cette étude, ce sont les algorithmes formés suivant les méthodes d'assemblage *Adaboost* (DT renforcés) et *Bagged trees* (DT ensachés ou *Random Forest*) qui ont été utilisés sur les données d'apprentissage. La flexibilité des modèles résultants augmente avec le paramètre « Nombre d'apprenants ». Tous les classifieurs d'ensembles ont tendance à être lents afin de s'adapter. Car, ils ont souvent besoin de beaucoup d'apprenants. Le paramètre « Nombre maximum de divisions » demeure pour ces algorithmes, le nombre maximal de fractionnement ou de points de branche à utiliser pour contrôler la profondeur des arbres.

De même que les classifieurs à arbre unique, les arbres trop profonds peuvent provoquer un surajustement sur les données d'apprentissage. En effet, des branches trop nombreuses peuvent avoir tendance à surajuster le modèle aux données d'entrées, tandis que des arbres plus simples peuvent être plus robustes et faciles à interpréter. Dans le but d'évaluer la capacité prédictive de différents types de classifieurs d'ensemble et d'obtenir des détails de prévision les plus fins et les plus optimaux pour les jeux de données, des modèles basés sur des arbres plus profonds et sur un plus grand nombre d'arbres peu profonds ont été construits et testés dans cette étude. Cela a permis d'obtenir une profondeur optimale pour l'apprentissage sur notre jeu de données et de trouver un compromis entre l'ajustement des

données et leur complexité. Pour se faire, plusieurs valeurs des paramètres « Nombre d'apprenants et Nombre maximal de divisions » ont été utilisées. En effet, un modèle d'ensemble d'arbres renforcés a une flexibilité élevée et celle-ci augmente également avec le paramètre « Nombre de divisions ». Il en est de même pour les modèles d'arbres ensachés ou RF qui présentent également une flexibilité très élevée.

Le paramètre « Taux d'apprentissage » est le paramètre qui fixe une condition d'arrêt de l'apprentissage. Un taux d'apprentissage inférieur à 1 nécessite plus d'itérations d'apprentissage, mais permet d'obtenir souvent une meilleure précision. Le taux de 0,1 est un choix beaucoup plus populaire et a été testé dans cette étude. Le taux d'apprentissage de 1 a également été testé sur le jeu de données d'apprentissage quand ce fut possible.

Enfin, le paramètre de « Dimension de sous-espace » spécifie le nombre de prédicteurs à échantillonner dans chaque apprenant. Ce paramètre permet à chaque apprenant de choisir de façon indépendante un sous-ensemble aléatoire de prédicteurs. Dans cette étude, la valeur 1 a été utilisée pour ce dernier paramètre.

3.8 Technique d'amélioration de performance d'un classifieur

Le concept de compromis biais-variance permet de trouver un équilibre entre un modèle surentraîné et un modèle sous-entraîné. Un modèle surentraîné est un modèle complexe de variance élevée, qui risque le surapprentissage (biais faible sur les données d'entraînement, mais élevé sur de nouvelles données). Un tel modèle aura de très bonnes performances sur le jeu d'entraînement, mais sera mauvais sur de nouvelles données. En revanche, modèle simple de variance faible risque le sous-apprentissage (biais élevé y compris sur les données d'entraînement). Pour satisfaire à ce concept, une première approche a été utilisée pour améliorer et évaluer les performances d'apprentissage de chaque classifieur. Cette technique a été appliquée au cours du processus d'apprentissage.

Elle a consisté à intégrer dans le processus, une étape de validation croisée. Cela permet d'éviter d'éventuels biais liés au phénomène de surajustement lors de l'apprentissage et d'évaluer la performance prédictive de chaque algorithme sur de nouveaux jeux de données, sur lesquels ils n'ont pas encore été formés. Cette technique donne également, une bonne estimation sur la précision prédictive du modèle final, formé avec toutes les

données d'entrées. Plusieurs techniques de validation croisée sont couramment utilisées dans la littérature en matière de CPM (Ortiz *et al.*, 2006; Peters *et al.*, 2007; Mahmoudi *et al.*, 2012; Toth *et al.*, 2013).

Dans cette étude, c'est la technique de « *Leave-one-out* » qui a été appliquée. Elle a consisté à partitionner les données d'apprentissage en utilisant l'approche des « *k-fold* », où k est égal au nombre total d'observations dans les données. Compte tenu de la petite taille de notre jeu de données initial (3709 points), un nombre k égal à 5 a été utilisé pour limiter les effets de surajustement lors de l'apprentissage. Chaque cycle de validation croisée implique la partition aléatoire de l'ensemble de données d'origine en un sous-ensemble d'apprentissage et un sous-ensemble test. Le sous-ensemble d'apprentissage est ensuite utilisé pour former l'algorithme d'apprentissage et le sous-ensemble test est utilisé pour évaluer ses performances. L'erreur de validation croisée moyenne est utilisée comme indicateur de performance. Pour chaque cycle de formation, ce processus est répété de la sorte cinq fois.

3.9 Technique d'évaluation de performance d'un classifieur

La technique utilisée pour évaluer la performance d'un classifieur sur un jeu de données d'apprentissage a consisté à examiner une série d'indices statistiques. Il s'agissait d'estimer la précision de classification globale et de comparer les valeurs d'AUC résultantes de chaque classification. Cette étape a permis de sélectionner le meilleur algorithme dans chaque catégorie (SVM, KNN et DT), à savoir le plus précis et le plus performant sur le jeu de données d'apprentissage. Les résultats de cette analyse sont trois meilleurs modèles entraînés SVM, KNN et DT. Par la suite, quatre approches comparatives croisées ont été appliquées sur ces meilleurs modèles entraînés, de sorte à identifier lequel est le plus précis, le plus performant et le plus robuste parmi les trois.

La première approche d'analyse comparative croisée a consisté à comparer les trois valeurs de précisions globales obtenues par les meilleurs modèles.

La deuxième approche a consisté à comparer les valeurs d'AUC obtenues pour chacune des classes de cibles minérales. Cette étape a permis de comparer la performance des meilleurs modèles sur un même jeu de données d'apprentissage et d'identifier le plus

performant en termes de prédiction de classes (cibles minérales). La précision globale d'une classification est un indice optimiste des performances d'un classifieur. Cependant, il importe de tenir compte dans l'évaluation des performances d'un modèle, des résultats réels de classification des entités proposées dans le jeu de données d'apprentissage. Car, les valeurs d'AUC et les précisions globales obtenues, sont des estimations de la performance globale des modèles. En revanche, l'analyse d'une matrice de confusion permet de mesurer de façon plus pertinente les performances réelles d'un modèle entraîné. En effet, les matrices de confusions fournissent un résumé détaillé des résultats de prédictions, sur le problème de classification posé. Les prédictions correctes et incorrectes y sont ainsi mises en lumière et réparties par classe. Les prédictions peuvent alors être comparées avec des valeurs réelles. En somme, il s'agit de vérifier à quelle fréquence les prédictions du modèle sont exactes par rapport à la réalité posée dans le problème de classification. De plus, l'analyse des matrices de confusions permet de comprendre de quelle façon un modèle de classification est confus lorsqu'il effectue ses prédictions. Cela permet, non seulement, de savoir quelles sont les erreurs commises, mais surtout de connaître le type d'erreurs commises.

Dans cette optique, une troisième étape d'analyse comparative a été réalisée, pour permettre d'évaluer la capacité prédictive réelle de chaque meilleur modèle (KNN, SVM et DT) sur les entités cibles minérales. Tout d'abord, la sensibilité et la spécificité de chaque meilleur modèle ont été évaluées à travers les courbes ROC (*Receiver Operating Characteristic*) et les valeurs d'AUC obtenue pour chaque classe de cible minérale. Les courbes ROC expriment la capacité de discrimination d'un modèle entraîné. Une allure convexe de la courbe ROC représente une capacité de discrimination du modèle entraîné avec une valeur d'AUC comprise entre 1 et 0,5. L'AUC fournit également une mesure agrégée des performances pour tous les seuils de classification possibles. On peut interpréter l'AUC comme une mesure de la probabilité pour que le modèle classe un exemple positif aléatoire au-dessus d'un exemple négatif aléatoire. Les valeurs d'AUC sont comprises dans une plage de 0 à 1. Un modèle dont 100 % des prédictions sont erronées présente un AUC de 0,0. Si toutes ses prédictions sont correctes, son AUC est de 1,0.

Par la suite, les Valeurs Prédicatives Négatives (VPN), les Valeurs Prédicatives Positives (VPP), les Taux de Vrais Positifs (TVP) et les Taux de Vrais Négatifs (TVN) issus des différentes matrices de confusions ont été comparés et analysés. Cette étape a permis d'évaluer la capacité prédictive réelle de chaque meilleur modèle et d'identifier lequel est véritablement le plus performant sur le jeu de donnée d'apprentissage. Le calcul de ces différents paramètres peut être effectué suivant les équations suivantes (Tien Bui *et al.*, 2016):

$$\text{Sensibilité ou TVP} = \frac{TP}{TP + FN} \quad (\text{Éq. 15})$$

$$\text{Spécificité} = \frac{TN}{TN + FP} \quad (\text{Éq. 16})$$

$$\text{Valeur Prédicative Positive} = \frac{TP}{TP + FP} \quad (\text{Éq. 17})$$

$$\text{Valeur Prédicative Négative} = \frac{TN}{TN + FN} \quad (\text{Éq. 18})$$

$$\text{Précision globale} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{Éq. 19})$$

Avec :

TP : le taux (exprimé en pourcentage) de vrais positifs (fais référence aux échantillons bien classés) ;

FN : le nombre de faux négatifs (fais référence aux échantillons positifs mal classés) ;

FP : le nombre de faux positifs (fais référence aux échantillons négatifs mal classés) ;

TN : le taux (exprimé en pourcentage) de vrais négatifs (fais référence aux échantillons négatifs bien classés).

Dans certains cas, une dissymétrie dans le jeu de données d'apprentissage peut conduire à une appréciation biaisée (surestimation ou sous-estimation) de la performance des modèles en termes de précision globale. En effet, le calcul de la valeur de précision globale tient compte de toutes les performances obtenues sur l'ensemble des classes présentes, d'où un risque de propagation des erreurs de commissions et d'omissions dans le calcul de la précision globale de classification. Dans ces cas, une technique souvent utilisée pour

obtenir une estimation réelle de la précision de classification, est l'utilisation de l'indice statistique F-score, également appelé score F1 ou mesure F. Le score F peut fournir une mesure plus réaliste de performance d'un test, en utilisant à la fois la précision et le rappel de chaque classe. Il s'agit tout simplement d'une mesure de précision d'un test effectué sur chacune des classes. Le score F est défini comme la moyenne harmonique pondérée de la précision et du rappel du test. Ce score est calculé selon l'équation suivante :

$$F1 = 2 \times \frac{(Précision \times rappel)}{(précision+rappel)} \quad (\text{Éq. 20}) \quad \text{avec } rappel = sensibilité \quad \frac{TP}{TP+VN} \quad (\text{Éq. 21})$$

La précision, également appelée valeur prédictive positive, est la proportion de résultats positifs réellement positifs. Le rappel, également appelé sensibilité, est la capacité d'un test à identifier correctement les résultats positifs, pour obtenir le taux de réponse positive réel. Le score F atteint la meilleure valeur, c'est-à-dire une précision et un rappel parfaits, à une valeur de 1. Le pire score F, ce qui signifie la plus faible précision et le plus faible rappel, correspond à 0. Le score F peut être alors utilisé pour mesurer la précision réelle d'une classification, car il équilibre l'utilisation de la précision et du rappel (Éq. 21), contrairement au calcul de la précision globale (Éq. 19).

Sur la base de l'analyse et de l'interprétation de ces différents résultats, les performances des trois meilleurs modèles seront évaluées. La prochaine étape a consisté à satisfaire aux derniers tests comparatifs en testant les performances et la robustesse des meilleurs modèles sur un jeu de données externes à leur apprentissage. Ce qui nous conduit à la présentation de la quatrième et dernière technique d'analyse comparative croisée.

3.10 Technique de validation et test de robustesse des modèles prédicteurs

La quatrième analyse comparative croisée a consisté à tester la robustesse de chaque meilleur modèle sur un jeu de donnée externe. En effet, obtenir de bonnes performances sur un jeu d'entraînement, ne garantis guère que le modèle soit assez robuste et capable de performer sur de nouvelles données, sur lesquelles il n'a pas été formé. Pour se faire, les données de validation (20 % du jeu de données initial) ont été soumises à chaque meilleur modèle. Les matrices de confusions ont été analysées, les valeurs d'AUC estimées et les coefficients kappa calculés et comparés.

Le coefficient Kappa est généré à partir d'un test statistique pour évaluer la précision d'une classification. Kappa évalue essentiellement la qualité de la classification effectuée par rapport à l'attribution aléatoire de valeurs, c'est-à-dire, évaluer si la classification a été meilleure que la statistique aléatoire. Le coefficient Kappa peut aller de -1 à +1. Une valeur de 0 indique que la classification n'est pas meilleure qu'une classification aléatoire. Un nombre négatif indique que la classification est significativement pire que le hasard. Une valeur proche de 1 indique que le classement est nettement meilleur que le hasard. Le calcul du coefficient kappa peut être effectué suivant l'équation suivante (Tien Bui *et al.*, 2016) :

$$\text{Coefficient Kappa de Cohen} = \frac{(TP+TN) - [(TP+FN)(TP+FP) + (FP+TN)(FN+TN)] / (TP+FP+FN+TN)}{(TP+FP+FN+TN) - [(TP+FN)(TP+FN)(TP+FP) + (FP+TN)(FN+TN)] / (TP+FP+FN+TN)} \quad (\text{Éq. 22})$$

Les résultats de cette étape ont permis de valider les performances des meilleurs classifieurs KNN, SVM et DT et de déterminer lequel des modèles prédictifs était le plus précis, le plus sensible, le plus spécifique, en termes de prédiction sur un jeu de données externes.

L'étape de modélisation prédictive se termine ainsi avec l'analyse des résultats de ces tests et la sélection du meilleur modèle prédictif. L'algorithme le plus précis, le plus performant et le plus robuste est le modèle prédictif final. Ce dernier a alors été utilisé pour cartographier dans un environnement SIG, les nouvelles zones prospectives potentielles de la zone d'étude.

3.11 Méthode de cartographie prédictive dans un SIG

La modélisation prédictive de la prospectivité minérale à l'aide d'un SIG est un outil valide et de plus en plus accepté pour délimiter des cibles d'exploration minérale reproductibles (Sun *et al.*, 2019). La plupart des approches analytiques dans un SIG, utilisent le concept de conversion de données vectorielles ponctuelles en données matricielles afin d'affecter à une cellule de dimension spécifique, l'information contenue dans une entité ponctuelle. Cela permet de cartographier, de visualiser et d'analyser de façon spatiale la variation d'un phénomène ponctuel sur une surface continue.

La partie nord de la zone d'étude est contenue dans un rectangle de 21,83 km de long sur 28,17 km de large. La partie sud de la zone d'étude est contenue dans un rectangle de 14,39

km de long sur 17,88 km de large. La méthode appliquée de cartographie prédictive minérale dans un environnement SIG, a consisté à subdiviser d'abord les deux parties (rectangles) de la zone d'étude selon deux grilles matricielles respectives à maille régulière. Une maille élémentaire de 25 m x 25 m a été utilisée en raison de la dimension de la maille prospective (100 m x 50 m) utilisée lors de la prospection des sites de Bémadi, Brokoua et Makoua, d'où sont issues les données de prospection minière de l'étude. La distance de 25 m correspond à la moitié de la distance minimale qui sépare deux puits voisins échantillonnés. Par la suite, les points « centroïdes » de chaque cellule des grilles matricielles ont été calculés. Au total, pour les deux grilles, 7 147 275 points ont ainsi été générés. Leurs coordonnées géographiques ont été calculées dans un SIG puis extraites dans une nouvelle table de données. La prochaine étape a consisté à superposer les cartes des prédicteurs spatiaux aux emplacements des nouveaux points centroïdes et à affecter les valeurs correspondantes aux coordonnées géographiques de ces points. La nouvelle table de données ainsi mise à jour a été utilisée dans l'outil de classification automatique comme nouveau jeu de données d'entrée. Ce jeu de données a servi alors à calculer et à prédire la réponse de classification (code des cibles minérales de 1 à 4) en utilisant (tour à tour) les trois meilleurs modèles prédicteurs (SVM, KNN et DT). Le résultat de ces prédictions est une nouvelle table de jeu de données, argumentée des codes de classes de cibles minérales prédites par ces modèles.

La dernière étape a consisté à appliquer une symbologie de couleur par type de cible minérale et à rastériser, suivant la maille de 25m x 25m, tous les points centroïdes. À cette étape, il convient de rappeler que dans un SIG, les stratégies d'attribution de valeurs et d'attributs ponctuels, à une unité surfacique élémentaire, obéissent à plusieurs règles d'affectation. Ces règles sont basées sur des calculs effectués sous certaines conditions de voisinage qui existent, entre l'emplacement du point à rastériser et ses plus proches voisins. En effet, lorsque la cellule de sortie ne contient qu'un seul point, la cellule prend la valeur et l'attribut de ce point. En revanche, lorsque la cellule contient plusieurs points et qu'aucune règle de priorité n'est définie ; ou lorsque les priorités sont identiques et que les attributs et les valeurs sont tous différents alors, l'attribut et la valeur associés à la plus petite cellule sont affectés à la cellule en sortie.

D'autre part, plusieurs règles prioritaires d'affectation d'attributs à une cellule en sortie peuvent être définies. Il s'agit soit de la règle de l'attribut le plus fréquent ou le plus commun, soit de la règle de calcul de déviation standard (entre la valeur du point à rastériser et ses voisines). Cependant, et dans tous les cas, si les règles prioritaires sont différentes et que les valeurs et les attributs sont différents, la priorité la plus élevée s'applique au champ prioritaire le plus commun. Enfin, en cas de doublon l'attribut associé au plus petit champ d'entrée est utilisé.

Un autre critère à prendre en considération lors de la rastérisation d'un point dans un SIG, est le critère de limite ou de bordure. En effet, lorsqu'une entité se trouve sur la limite d'une cellule, plusieurs règles de limite s'appliquent. Dans ce cas, les valeurs et les attributs des cellules situées en haut, en bas à gauche ou à droite de la limite, peuvent être définis comme prioritaires.

Dans le cadre de cette étude, la règle d'affectation de la valeur la plus fréquente, c'est-à-dire la valeur codée de classe de cible minérale la plus commune observée parmi des cellules voisines, a été appliquée pour les points situés en bordure de la grille à rastériser. Les valeurs et attributs des points situés à l'intérieur (en bas, en haut, à gauche et à droite) des bordures ont été priorisés.

La mise en œuvre de cette technique a ainsi permis de réaliser deux cartes prédictives de nouvelles zones potentiellement prospectives dans les parties nord et centre du Complexe géologique Hana-Lobo.

CHAPITRE IV. ANALYSES ET INTERPRÉTATIONS DES RÉSULTATS

Ce chapitre présente, analyse et interprète les différents résultats de l'étude.

4.1 Discrimination des différents types de cibles minérales

Les résultats de l'analyse statistique d'autocorrélation spatiale, consignés dans le tableau 3, montrent que la distribution géostatistique de certains groupes de graviers minéralisés sur les sites d'études n'est pas le fait d'un processus aléatoire. En effet, parmi les 3709 entités de graviers prospectés, 2920 peuvent être regroupées suivant quatre types de grappes spatiales dénommées *High High* : (HH), *High Low* : (HL), *Low Low* : (LL), *Low High* : (LH)). Il s'agit de 79 % des puits prospectés. Le reste des entités (21 %) ayant rejeté l'hypothèse nulle, ne présente pas d'agrégation spatiale statistiquement significative. Leur mode de distribution est par conséquent aléatoire. En effet, les résultats du tableau 3 montrent qu'à l'exception des 789 entités de la classe (AL) (dont les valeurs p sont supérieures à 0,04), toutes les valeurs p calculées pour les entités des grappes de type HH, HL, LL et LH, varient entre 0,002 et 0,038. Ces valeurs sont inférieures à 0,04. Cela signifie qu'il existe 2920 entités avec une agrégation spatiale statistiquement significative à 96 %.

Tableau 3 : Résultats de l'analyse statistique de Moran

Cotype	Nombre de points de grappe spatiale	Tranche de teneur de grappe spatiale (g/m ³)	Teneur moyenne en coltan par grappe spatiale (g/m ³)	Indice de Moran		Probabilité P (P-Value)		Nombre de voisins de grappe spatiale	
				Valeur		Valeur		Valeur	
				Minimale	Maximale	Minimale	Maximale	Minimale	Maximale
HH	797	45 à 1157	136,61	0,0005	15,35	0,002	0,038	7	554
HL	98	45 à 821	74,08	-1,65	0	0,002	0,034	33	442
LH	513	0 à 44	23,5	-0,59	0	0,002	0,038	9	546
LL	1 542	0 à 44	5,63	0	0,43	0,002	0,038	13	459
AL	789	0 à 1 126	40,98	-0,08	2,34	0,04	0,498	1	545
Total	3 709								

La figure 18 présente la courbe de répartition normale standard, associant les valeurs de scores Z et les valeurs p obtenues.

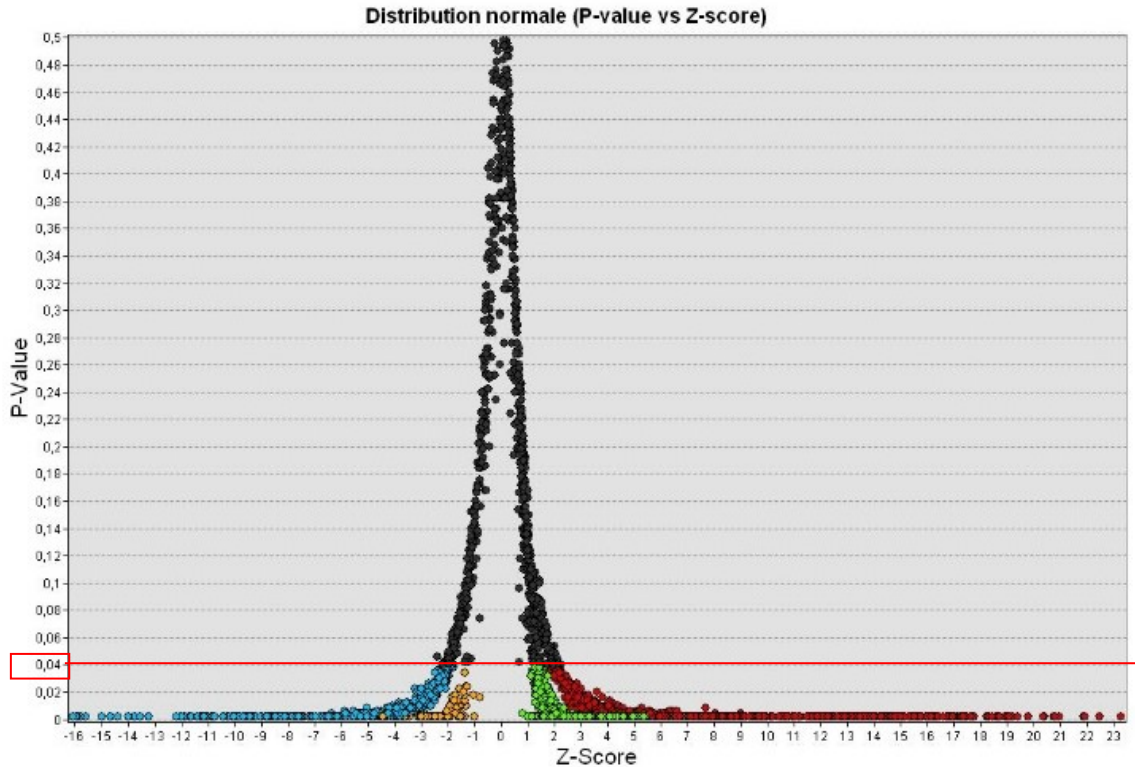


Figure 18 : Courbe d'autocorrélation spatiale (P-value vs Z-score.)

L'interprétation de la courbe montre que les scores z positifs ou très élevés (en rouge) et les scores z négatifs ou très faibles (en bleu) sont associés à des valeurs p très basses (inférieures à 0,04). Les valeurs des scores Z varient de +/- 2 à +/- 16. Ces deux types de grappes se trouvent aux extrémités de la courbe de répartition normale. De même, les scores Z des grappes de points aberrants (en orange et vert) varient de +/- 0,8 à +/- 5,5 et sont associées à des valeurs p inférieures à 0,04. Cela indique qu'il est peu probable que le modèle spatial observé (au niveau des quatre types de grappes) reflète un modèle aléatoire théorique. Cela indique que les points composant les grappes de couleur rouge, verte, bleue et orange présentent une agrégation spatiale statistiquement significative à 96 %. Il est donc, fortement improbable que le modèle spatial de distribution observé soit le résultat d'un processus aléatoire. En revanche, les points de couleur grise sont associés à des valeurs p supérieures à 0,04 et leurs scores z varient de 0 à +/- 3.

Ces points rejettent donc l'hypothèse nulle et ne présentent aucune agrégation spatiale statistiquement significative à 96 %.

L'analyse des valeurs de teneurs moyennes en coltan de ces différents types de grappes identifiées permet de tirer le constat suivant. Une valeur de 44 g/m³ peut être utilisée comme seuil de distinction des différents types de cibles minérales en présence. Le type de grappe de points (LL) représente des dépôts de graviers minéralisés à teneurs très faibles (en moyenne 5,63 g/m³ et inférieures à 44 g/m³). Le type de grappe de points (HH) représente des dépôts de graviers minéralisés à très fortes teneurs (en moyenne 136,61 g/m³ et supérieures à 44 g/m³). Le type (LH) représente des dépôts de graviers minéralisés à teneurs moyennes (en moyenne 23,5 g/m³ et inférieures à 44 g/m³) situées en périphérie des zones de dépôts de type (HH). Le type de grappe de points aberrants (HL) représente des dépôts de graviers minéralisés à teneurs fortes (en moyenne 74,08 g/m³ et supérieures à 44 g/m³) entourées de dépôts de type (LL).

En se basant sur ces différents modes d'agrégation spatiale et sur les valeurs des teneurs en coltan, les grappes de type HH, HL, LL et LH peuvent être utilisés pour représenter et caractériser les différents types de cibles minérales prospectives dans une CPM. D'un point de vue géoéconomique et dans le cadre des opérations de recherche minière à l'échelle régionale, ces différentes surfaces occupées par des dépôts minéralisés connus et avérés peuvent être utilisées à la fois pour modéliser de façon prédictive, en fonction de critères spatiaux de preuves, la prospectivité minérale de nouvelles régions d'intérêt.

D'autre part, les entités de la classe (AL) montrent une dispersion spatiale aléatoire et très peu de similarité entre les valeurs des teneurs moyennes de la classe (0 à 1126 g/m³) sur une superficie donnée.

Cependant, ces zones très peu prospectives pour des dépôts régionaux n'en sont pas moins stériles pour autant. À une échelle d'investigation beaucoup plus détaillée, à savoir à la dimension d'un puits minéralisé, ces surfaces pourraient dépendamment des intérêts stratégiques et économiques, représenter un certain intérêt. Pour ainsi dire, elles représentent statistiquement des surfaces où la probabilité de découvrir un dépôt de type HH, HL, LL, LH est inférieure à 4 %. Dans ces conditions et pour ces raisons, ces zones

ont été conservées dans la modélisation prédictive, afin de représenter les zones peu prospectives pour la découverte de dépôts de type HH, HL, LL, LH à l'échelle régionale. Tenant compte de ces interprétations et de ces résultats, quatre classes de cibles minérales prospectives ont été définies pour la modélisation prédictive.

La classe 1 (AL) est définie comme zone peu prospective, la classe 2 (LL) comme zone prospective à très faible teneur, la classe 3 (LH) comme zone prospective à teneur moyenne. La classe 4 est une classe mixte regroupant les grappes de type (HH) et (HL) à teneurs fortes à très fortes. Elle a été définie comme la classe cible de zone prospective pour des dépôts à teneurs élevées.

Compte tenu de l'insuffisance du nombre de puits réalisés dans la partie sud du Complexe géologique, l'analyse d'autocorrélation n'a pu être effectuée pour le site de Grand Zatry et Oureyo.

Les figures 19, 20 et 21 présentent respectivement les résultats cartographiques de l'analyse statistique d'autocorrélation spatiale réalisée pour les sites de Bémadi, Brokoua et Makua de la partie nord du Complexe géologique Hana-Lobo.

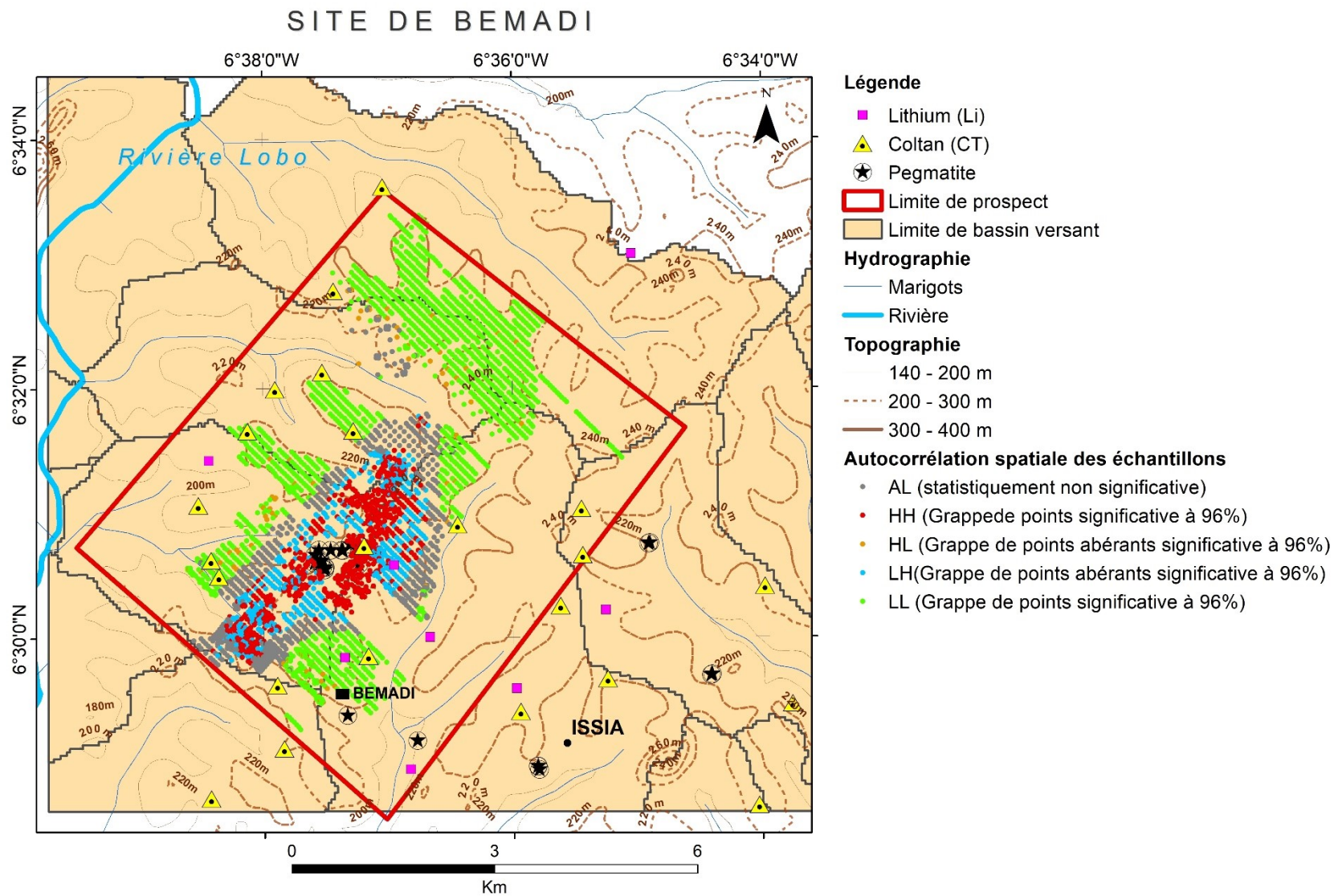


Figure 19: Carte des résultats de l'analyse statistique d'autocorrélation spatiale (Site de Bémadi)

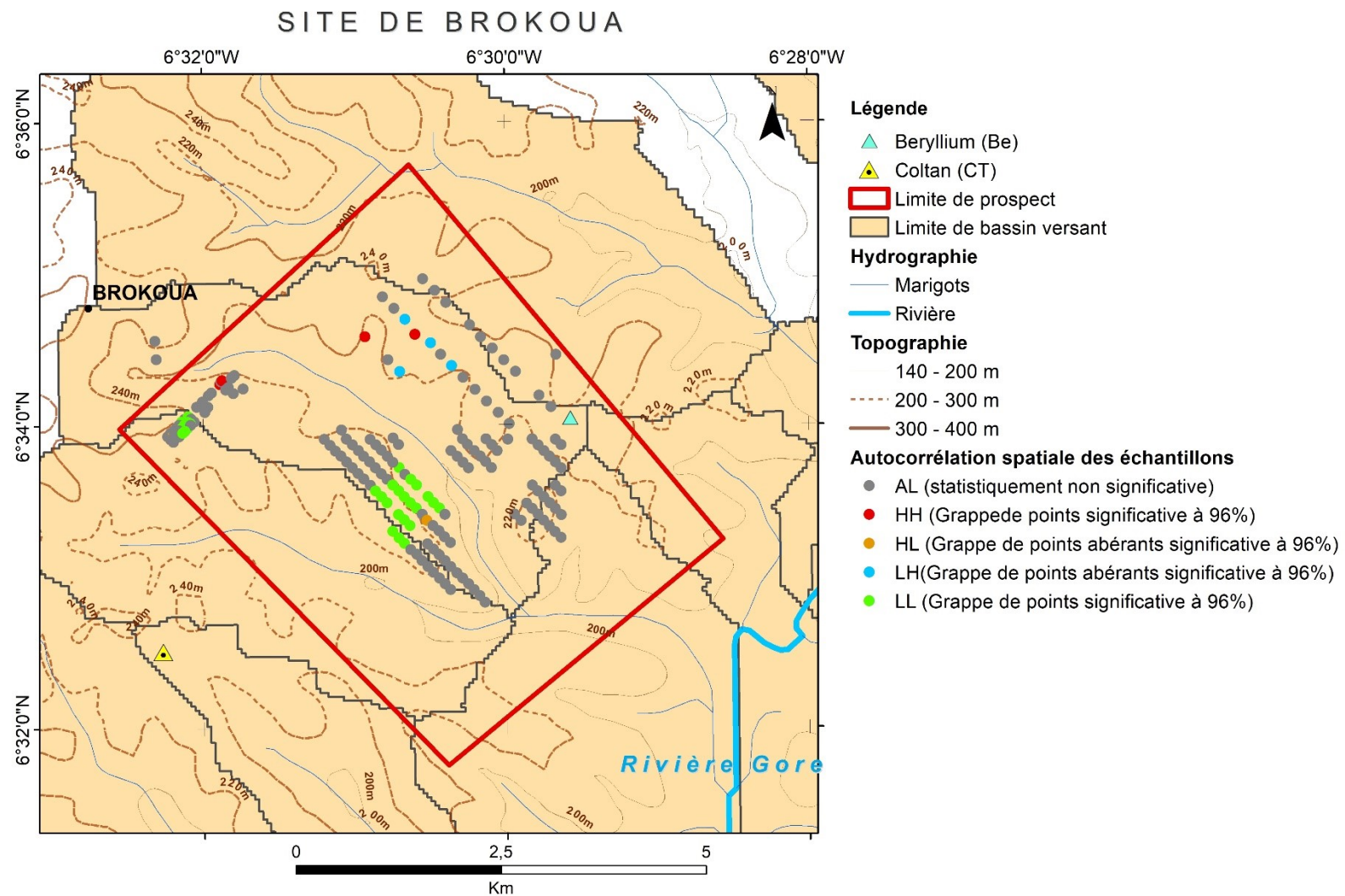


Figure 20 : Carte des résultats de l'analyse statistique d'autocorrélation spatiale (Site de Brokoua)

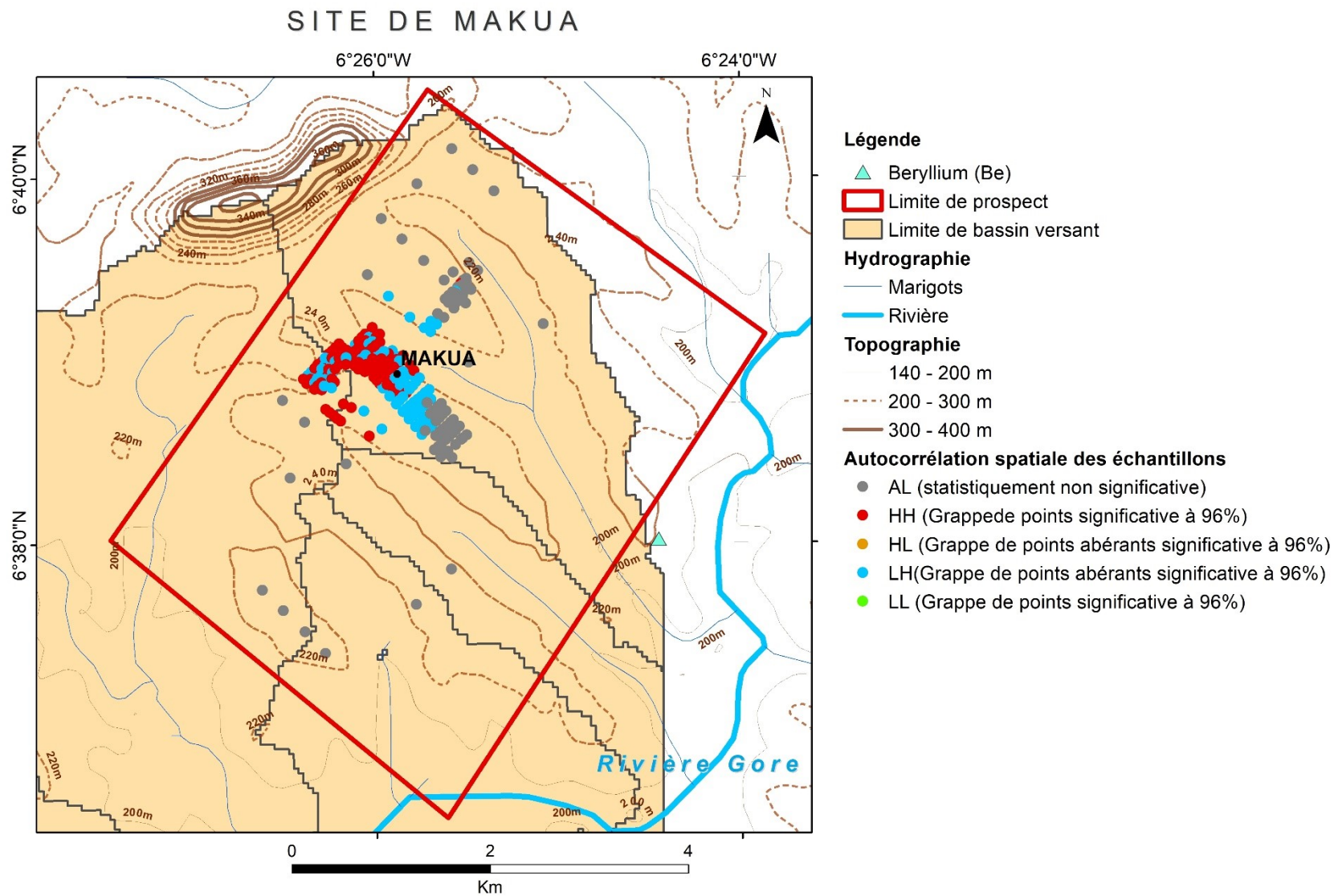


Figure 21: Carte des résultats de l'analyse statistique d'autocorrélation spatiale (Site de Makua)

4.2 Validation des linéaments naturels et analyse des prédicteurs spatiaux

Tout d'abord, il a fallu valider les linéaments naturels extraits de façon automatique. La figure 22 présente les linéaments naturels du site de Bémadi.

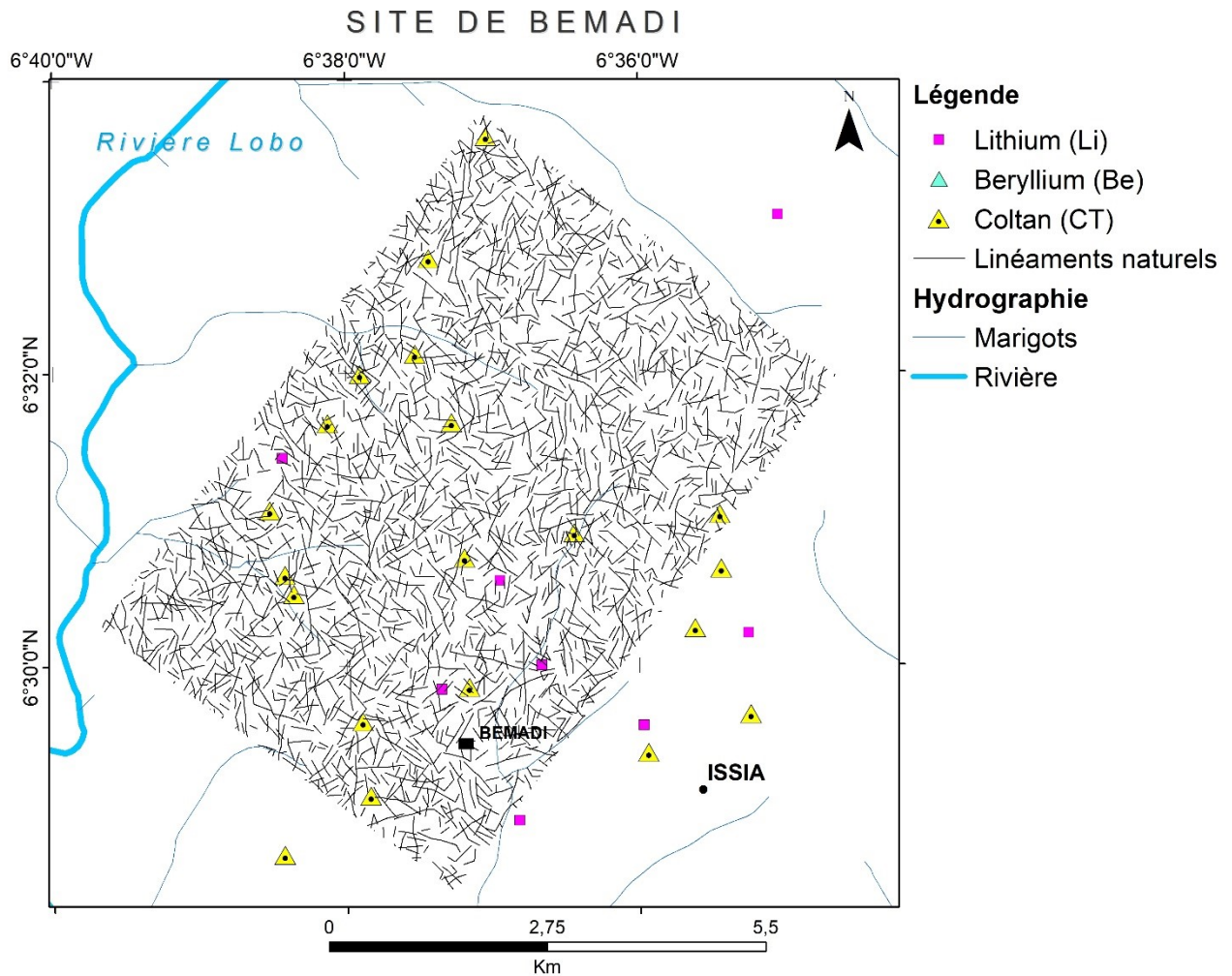


Figure 22: Carte des linéaments naturels du site de Bémadi

La figure 23 présente les linéaments naturels du site de Brokoua.

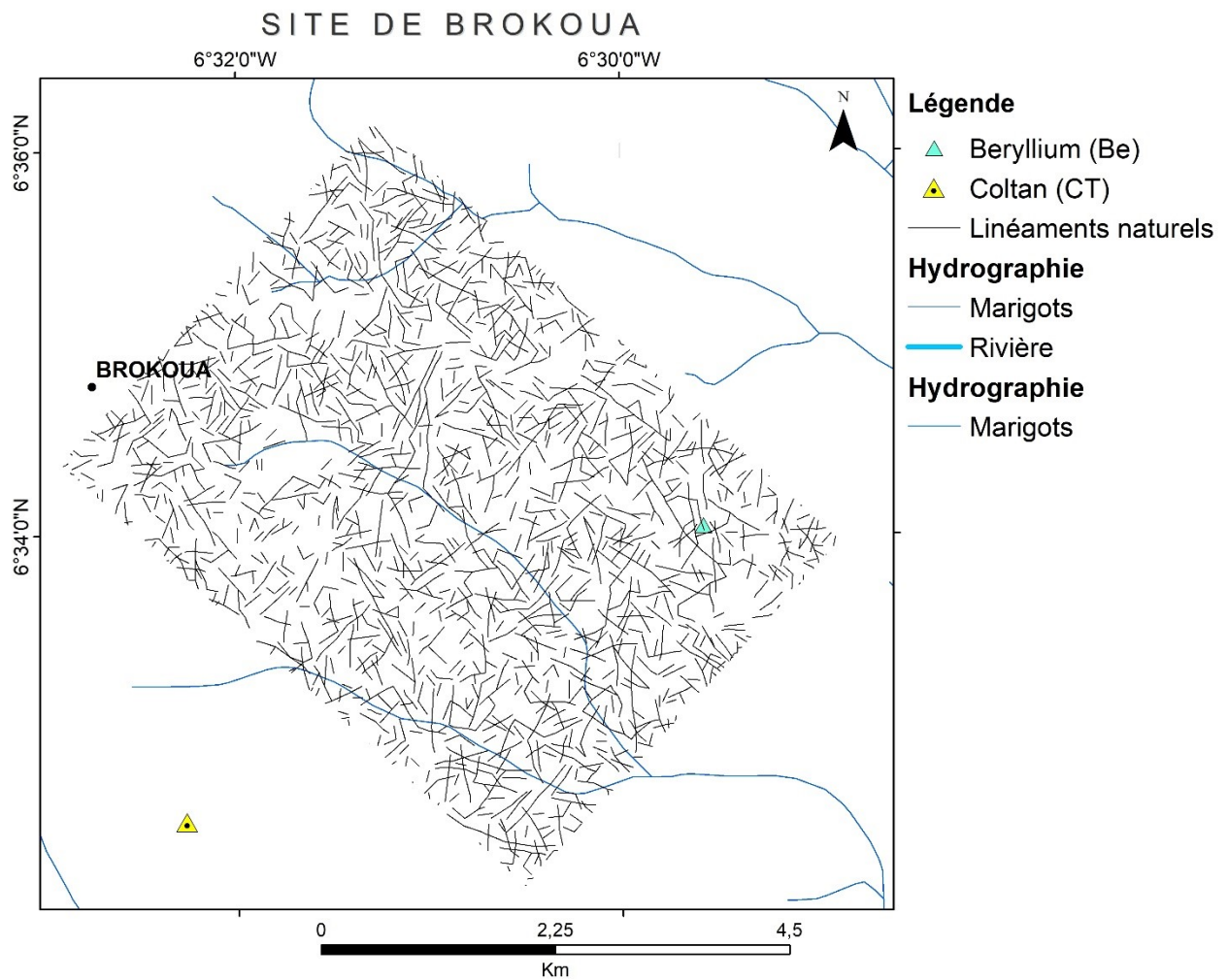


Figure 23 : Carte des linéaments naturels du site de Brokoua

La figure 24 présente les linéaments naturels du site de Makua.

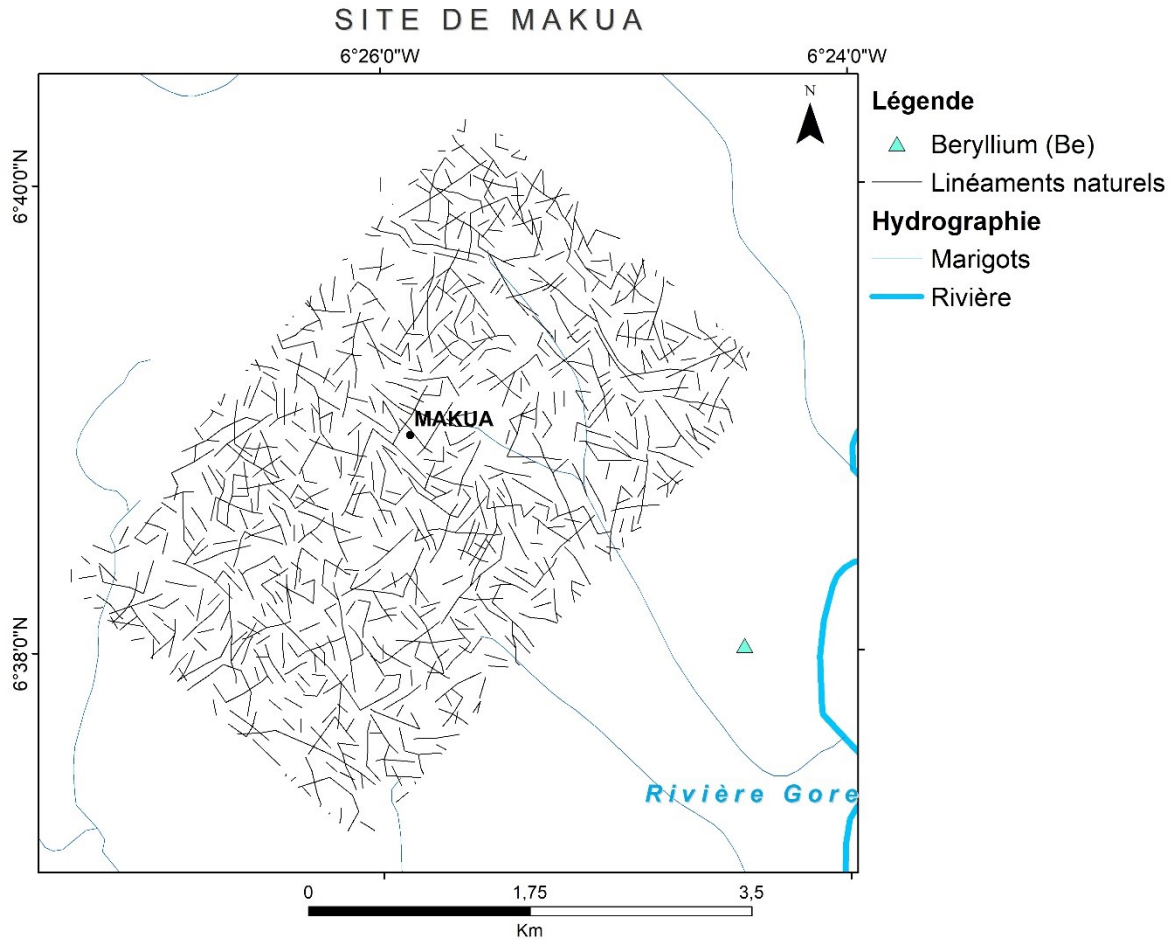


Figure 24: Carte des linéaments naturels du site de Makua

L'analyse et l'interprétation des directions préférentielles confirment l'existence d'un réseau de fracturations d'orientations très variées. Les directions de fractures majeures obtenues dans cette étude sont orientées NS (NN0-10), NE-SW (NN20-50), EW (NN90-110) et SE-NW (NN120-150), avec une direction moyenne régionale de 65,8 degrés. Ces directions observées et mises en évidence dans l'étude sont le résultat de contraintes majeures ayant affecté (au cours des différents stades orogéniques B, C et D) le socle précambrien et les formations post-éburnéennes du centre-ouest de la Côte d'Ivoire. Il s'agit de contraintes de cisaillement de direction principale (NS à NNW-SSE) du stade B,

des couloirs de décrochement du stade C et de la déformation transcurrente D2 (stade D) accompagnée par plusieurs fractures conjuguées (les fractures de Riedel et les fentes de tractions) (Allou, 2005). La figure 25 présente la rosace directionnelle obtenue.

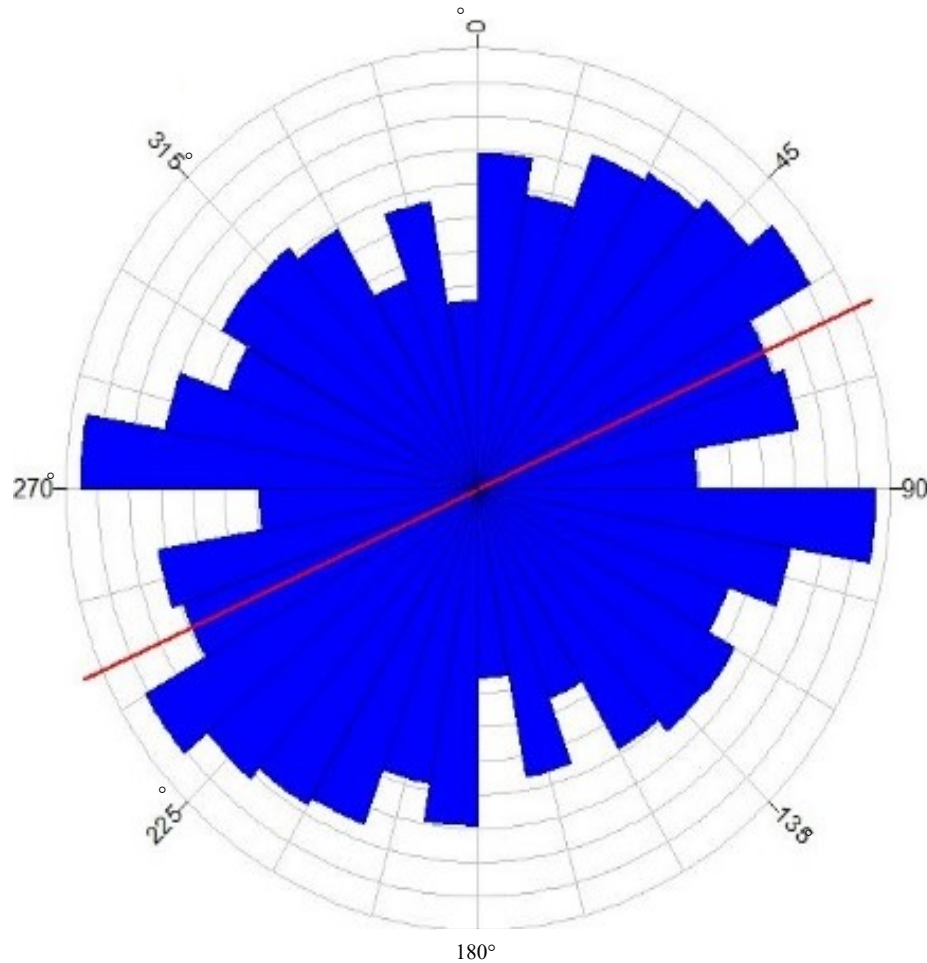


Figure 25 : Rosace directionnelle régionale.

Ces résultats sont en accord avec les directions régionales identifiées dans des travaux de cartographie structurale réalisés au centre-ouest de la Côte d'Ivoire (Koudou *et al.*, 2010, 2014) et plus particulièrement au niveau du bassin de la Lobo (Yao *et al.*, 2016).

Enfin, du point de vue morphogénétique, les directions mises en évidence dans cette étude, s'intègrent bien au système de déformation tectonique régional ayant donné naissance aux

plis droits et isoclinaux de la zone de contact schistes-granites, mentionnée par Triboulet *et al.* (1998) et Vidal *et al.* (2009). Par conséquent, les fractures et les failles extraites ont été utilisées pour cartographier et représenter de façon objective la variation spatiale de la densité de fracturation au niveau de la zone d'étude. Les résultats de l'étape d'extraction de critères de preuves sont des cartes de prédicteurs spatiaux à l'échelle 1/30 000 et présentées dans les figures 26 à 33. Ces cartes montrent la variation spatiale des valeurs des paramètres hydromorphologiques et structuraux utilisés comme prédicteurs spatiaux.

Paramètres de fracturation : Les cartes des figures 26 et 27 montrent la distribution des différents types de cibles minérales (HH, HL, LH, LL et AL) identifiées, en fonction de la densité de fracturation, respectivement sur le site de Bémadi et Brokoua - Makua.

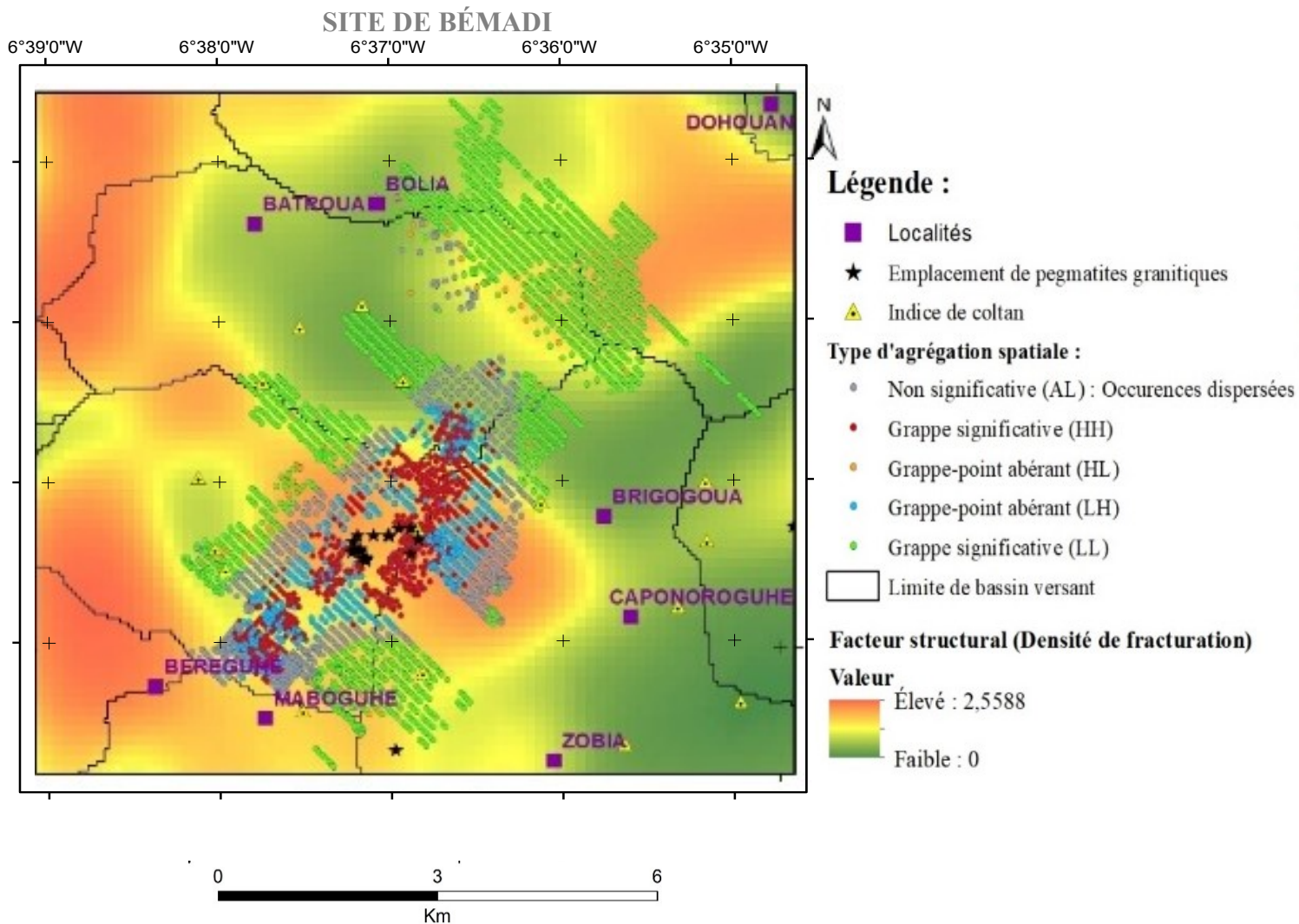
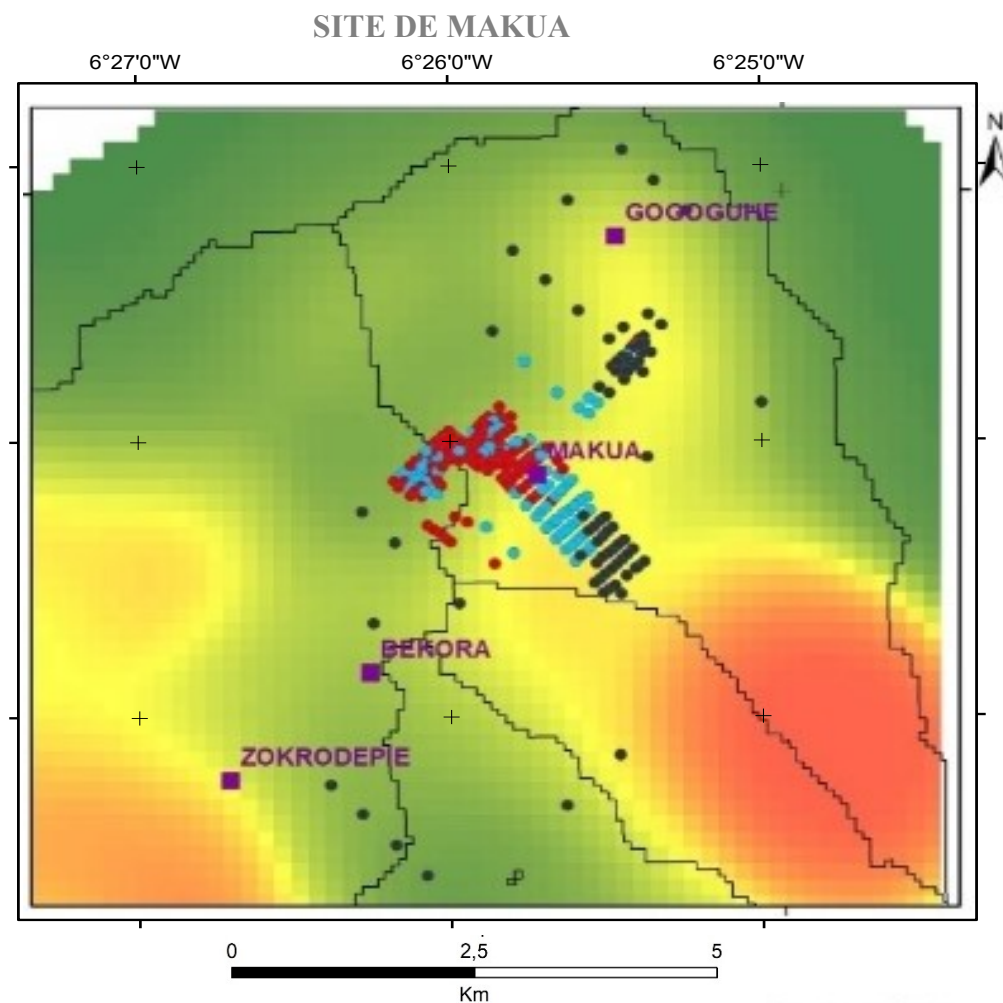
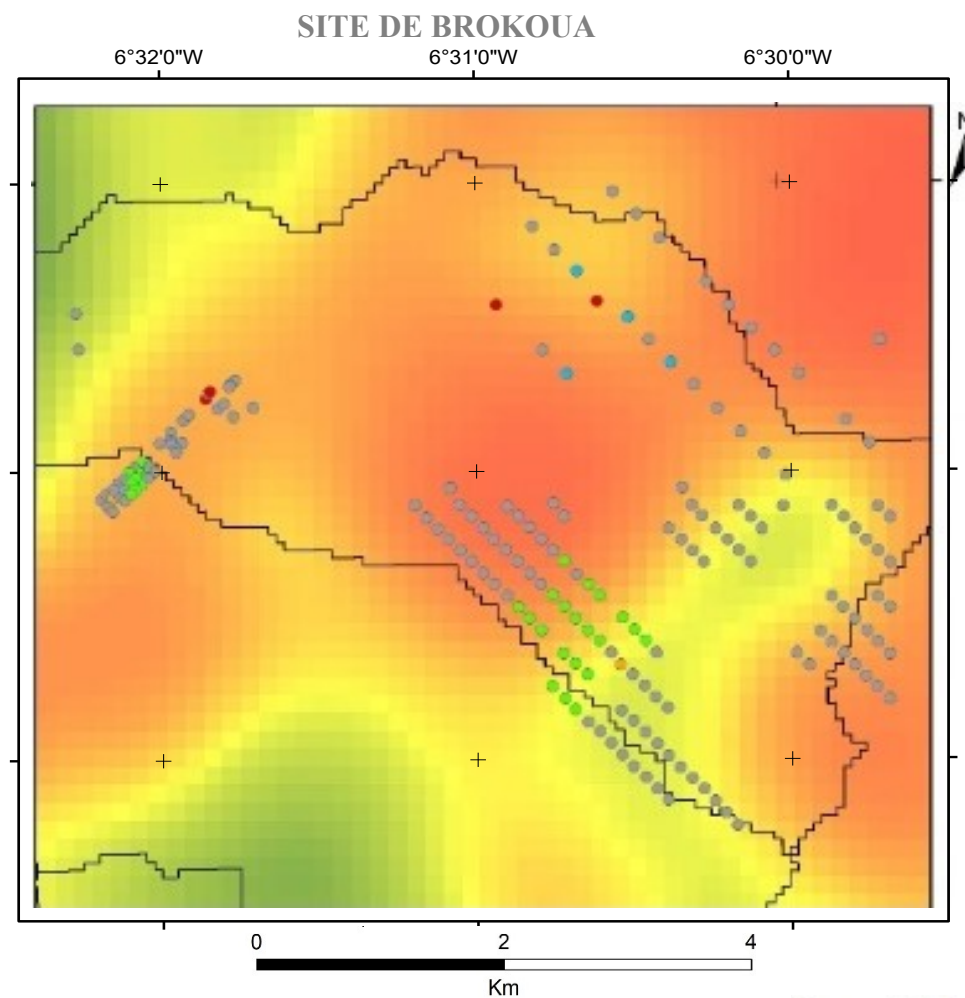


Figure 26: Cartes de densité de fracturation du site de Bémadi.



Légende :

- Localités
- Limite de bassin versant

Type d'agrégation spatiale :

- significative (AL) : Occurences dispersées
- type significative (HH)
- Grappe-point abérant (HL)
- Grappe-point abérant (LH)
- Grappe significative (LL)

Facteur structural (Densité de fracturation)

- Valeur**
- Élevé : 2,5588
 - Faible : 0

Figure 27: Carte de densité de fracturation des sites de Brokoua et Makua

L'observation de ces cartes montre que les valeurs du paramètre de densité augmentent dans la gamme de couleurs vert à rouge. Les zones de couleur jaune à jaune orangé représentent des zones de fracturation moyenne, situées à cheval entre des zones très fracturées en rouge et des zones moins fracturées en vert. L'interprétation de ces cartes montre que les emplacements d'affleurements connus de pegmatites granitiques minéralisés (particulièrement au niveau du site de Bémadi) coïncident avec une zone à forte densité de fracturation (en rouge). La plupart des cibles minérales de la classe HH, LH, LL et HL se situe, soit au cœur (site de Bémadi) ou à la périphérie (site de Makua et Brokoua) de zones fortement fracturées.

Paramètres morphologiques : Les figures 28 et 29 présentent la variation du paramètre de courbure (ou profil) de pentes respectivement sur les sites de Bémadi, Brokoua et Makua.

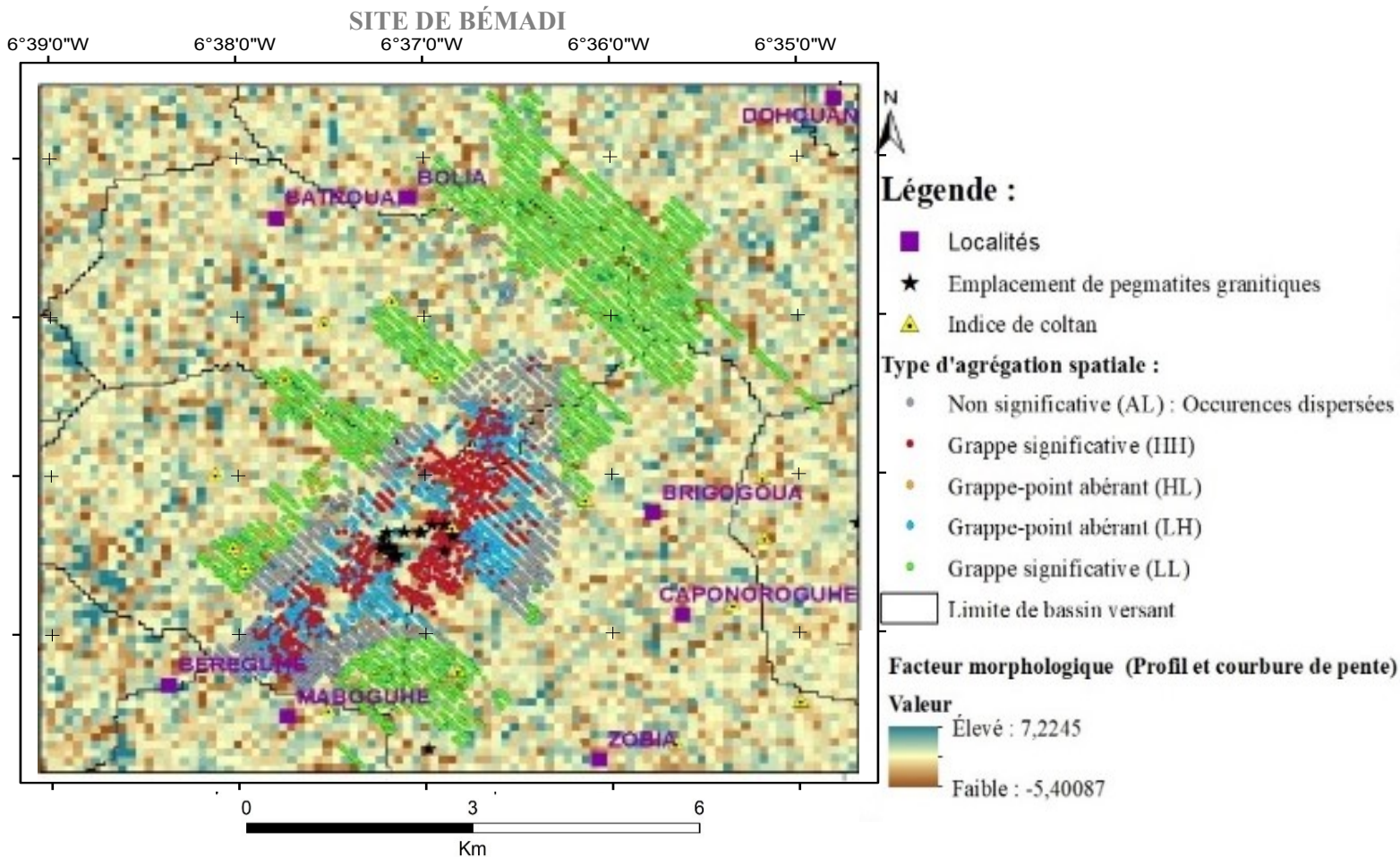
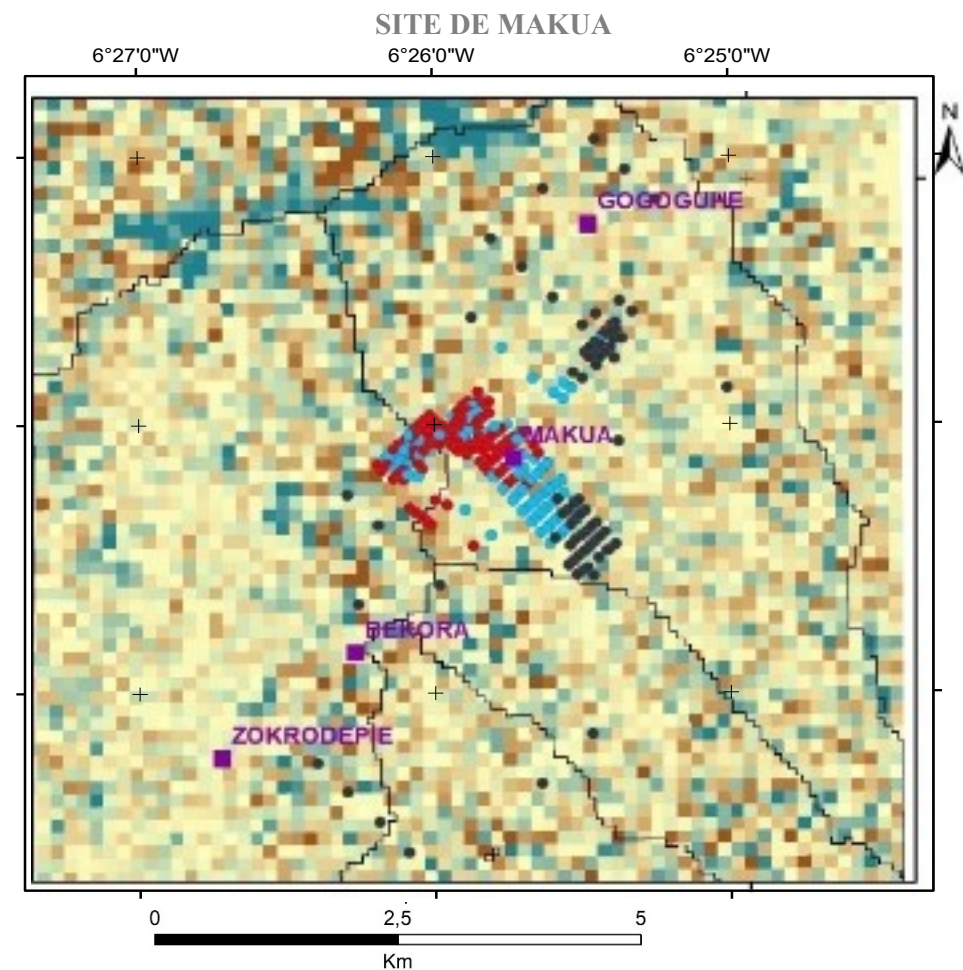
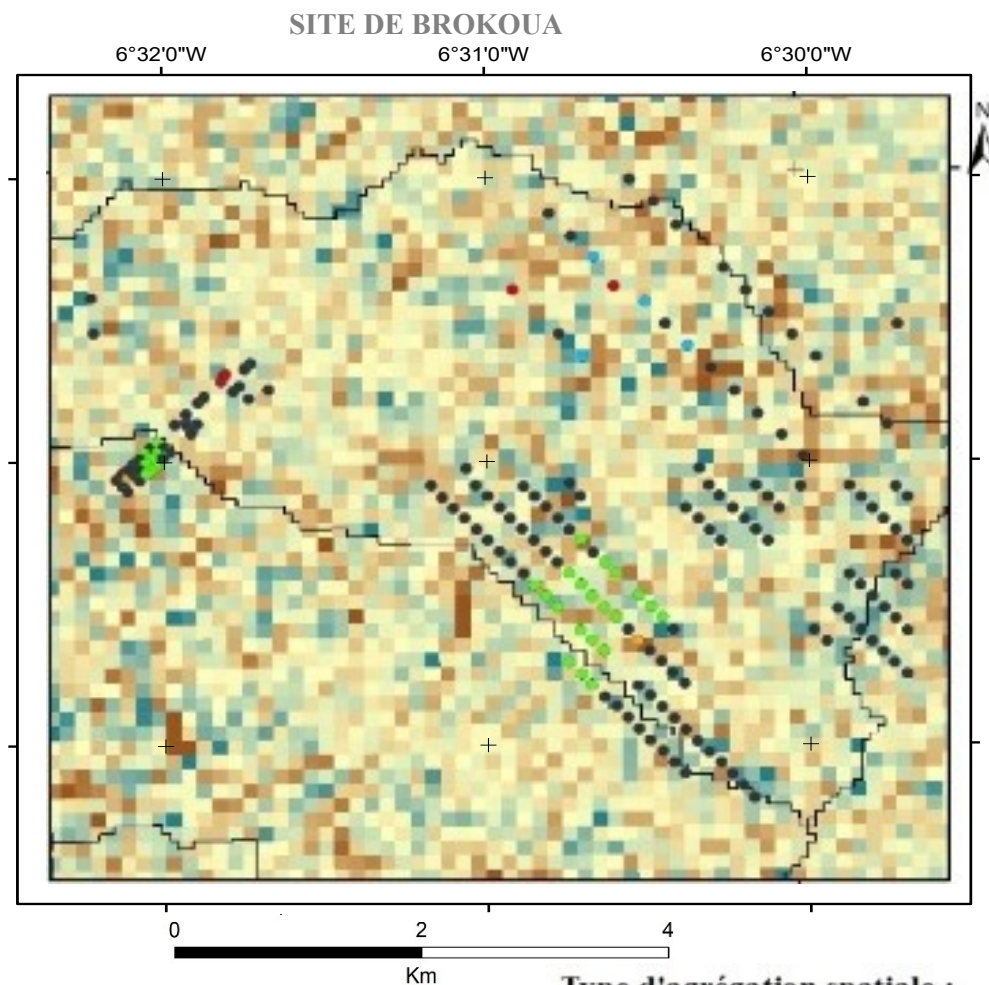


Figure 28: Cartes des profils et courbures de pentes du site de Bémadi



Légende :

- Localités
- Limite de bassin versant

Type d'agrégation spatiale :

- Non significative (AL) : Occurences dispersées
- Grappe significative (HH)
- Grappe-point abérant (HL)
- Grappe-point abérant (LH)
- Grappe significative (LL)

Facteur morphologique (Profil et courbure de pente)

- Valeur**
- Élevé : 7,2245
 - Faible : -5,40087

Figure 29 : Cartes des profils et courbures de pentes des sites de Brokoua et Makua

L'observation de ces cartes montre que les valeurs de courbures de pente augmentent dans la gamme de couleur marron foncé à bleu foncé. L'analyse de la carte de profil ou de courbure de pente met en évidence les zones de pentes très concaves (en marron foncé), favorables à l'érosion et les zones de pentes très convexes (en bleu foncé), favorables à l'écoulement des eaux de ruissellement.

Les zones quasi planiforme (en marron clair) et peu favorables à l'érosion des sols et à l'écoulement des eaux sont également mises en évidence sur ces cartes. Plusieurs zones de concentration de pixels de couleurs bleu foncé et marron foncé peuvent être observées. Ces zones sont situées en majorité en périphérie de limites de sous bassins versants (ligne noire sur les cartes) et sont généralement moins étalées que les zones quasi planiforme qui occupent la grande majorité des surfaces des sous-bassins versants. Cette variation illustre bien et confirme la présence d'irrégularités topographiques, observées dans les placers et mentionnées par plusieurs auteurs (Bouladon, 1964; Allou, 2005).

Cependant, l'interprétation de la distribution spatiale des différents types de cibles minérales (de dimension kilométrique) ne peut être uniquement expliquée, en fonction de ce seul paramètre de dimension métrique. En effet, aux plans morphologiques et topographiques, si la courbure et l'inclinaison des pentes donnent un aperçu cartographique des irrégularités topographiques à une échelle locale (échelle de 30 m), les cartes de relief en donnent un aperçu beaucoup plus régional (à une échelle kilométrique).

La figure 30 présente la variation du paramètre de l'inclinaison des pentes sur le site de Bémadi.

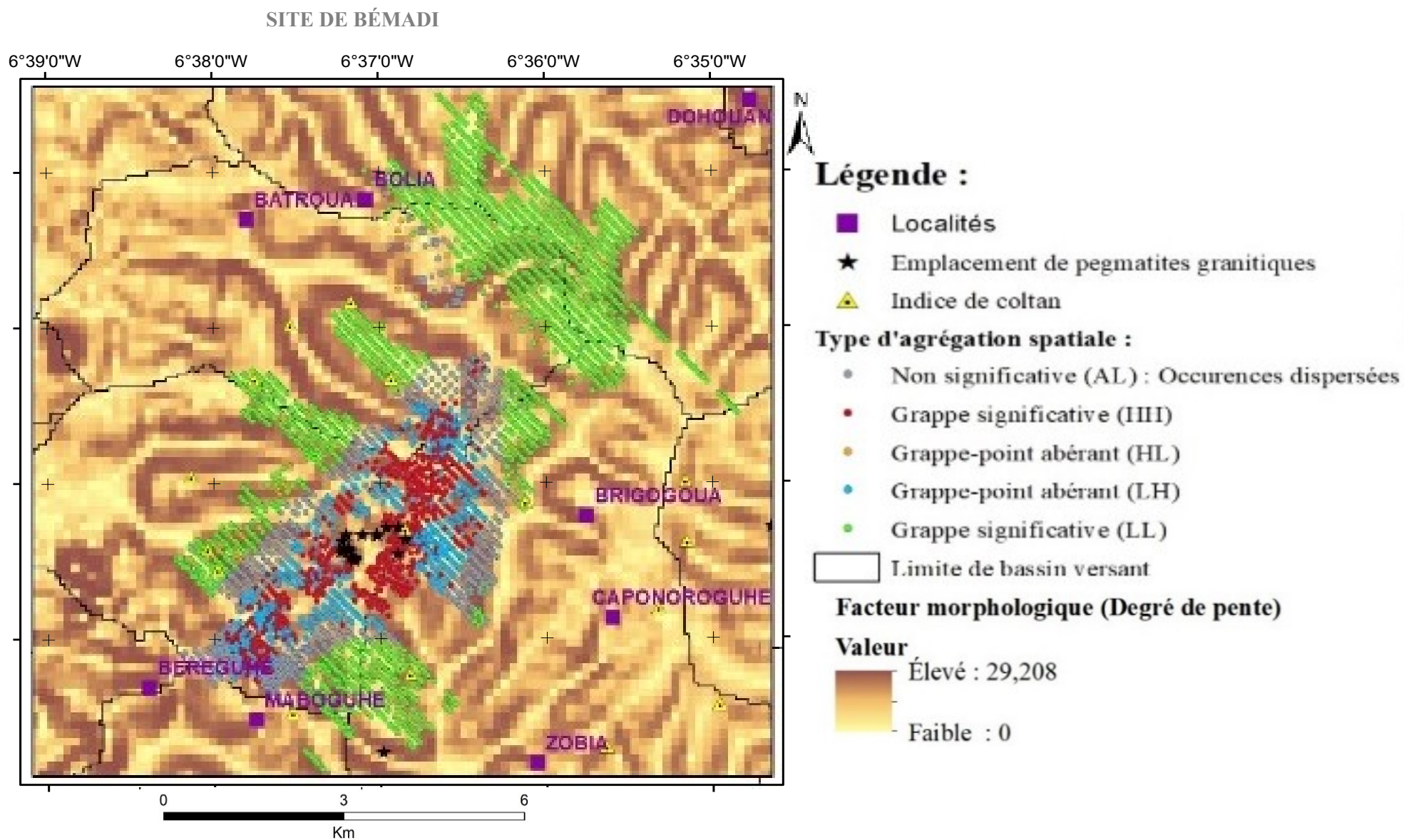


Figure 30: Cartes des degrés d'inclinaison des pentes du site de Bémadi

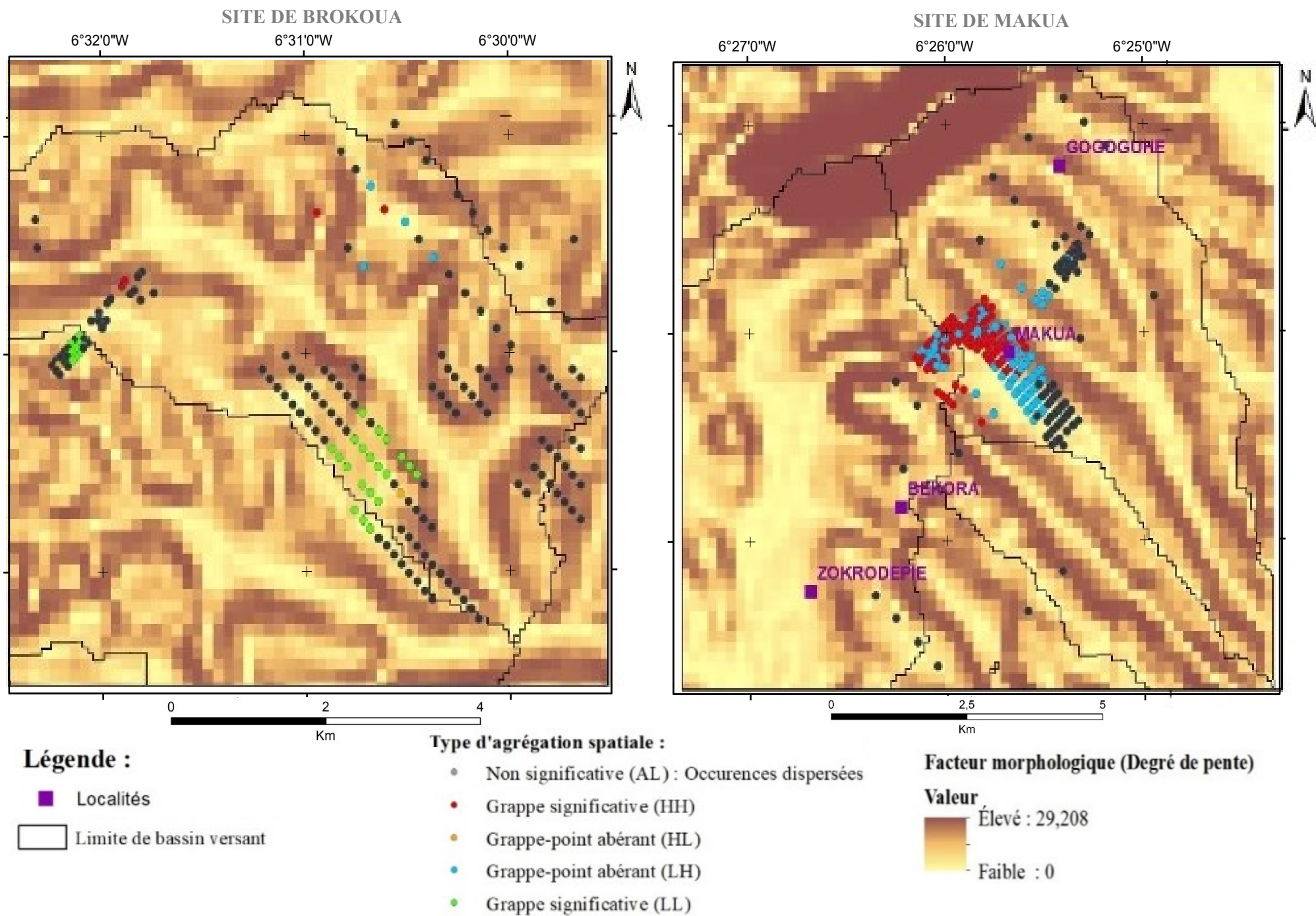


Figure 31: Cartes des degrés d'inclinaison des pentes des sites de Brokoua et Makua

L'observation des cartes montre que les valeurs de l'inclinaison des pentes augmentent dans la gamme de couleurs jaune clair à marron foncé. L'analyse et l'interprétation des cartes situent la grande majorité des surfaces occupées par des cibles minérales, au niveau des zones de pentes moyennes à faibles (marron clair). Cette observation traduit bien le fait que les pentes abruptes (en marron foncé) favorisent les processus d'écoulement et de drainage, tandis que les pentes de plus en plus douces contribuent à la décélération des eaux de ruissellement et donc, au piégeage et au dépôt des sédiments, au niveau des irrégularités topographiques. Cette observation traduit bien le fait que les pentes abruptes favorisent les processus d'écoulement et de drainage, tandis que les pentes de plus en plus douces contribuent à la décélération des eaux de ruissellement et donc, au piégeage et au dépôt des sédiments, au niveau des irrégularités topographiques. La figure 32 présente la variation du relief sur le site de Bémadi.

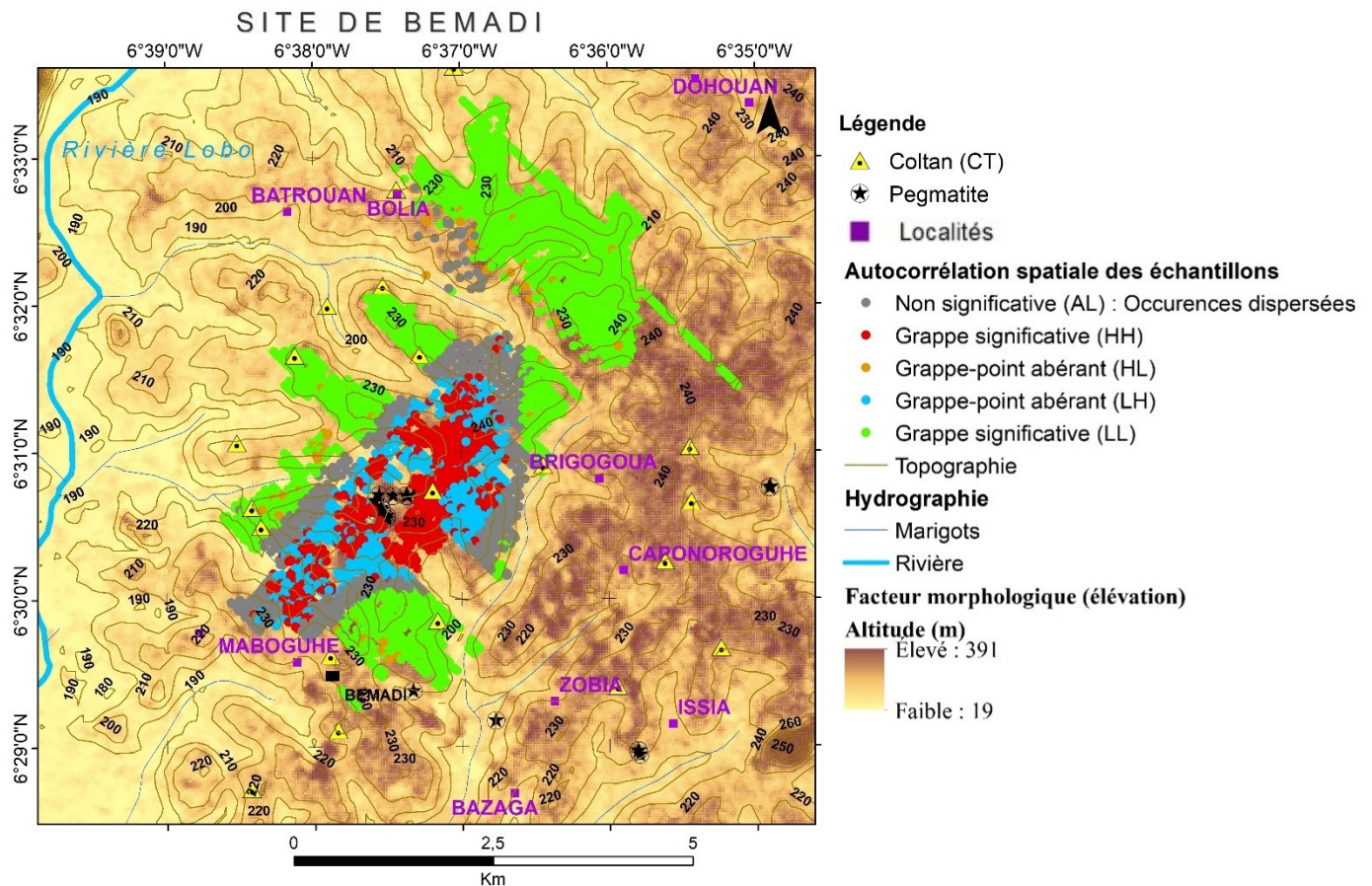
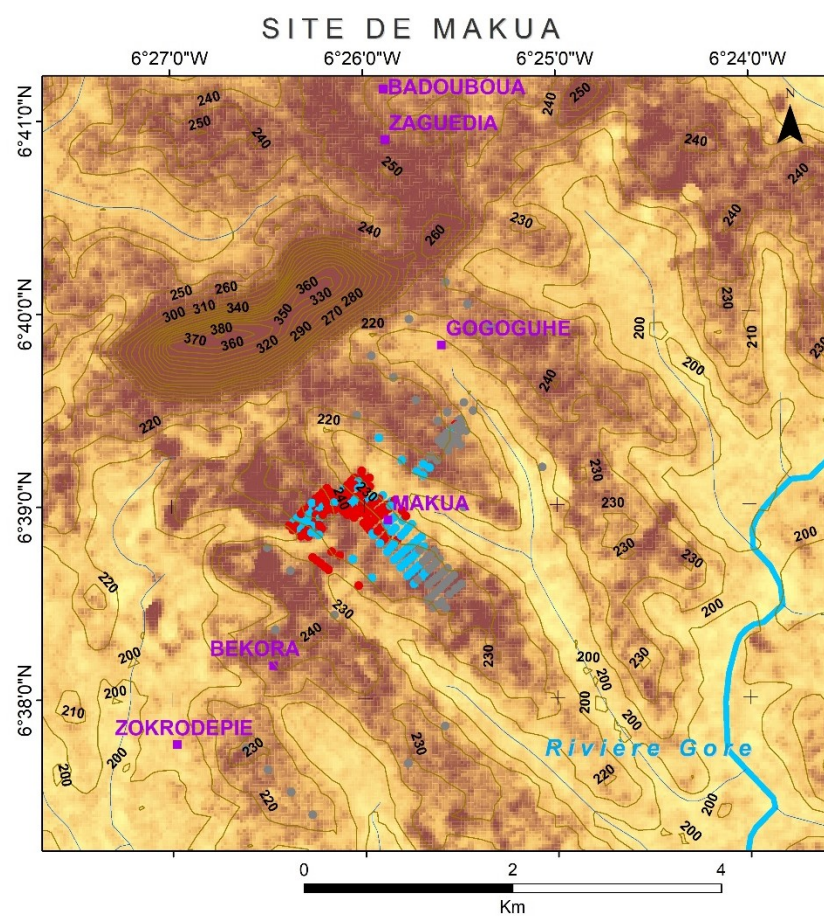
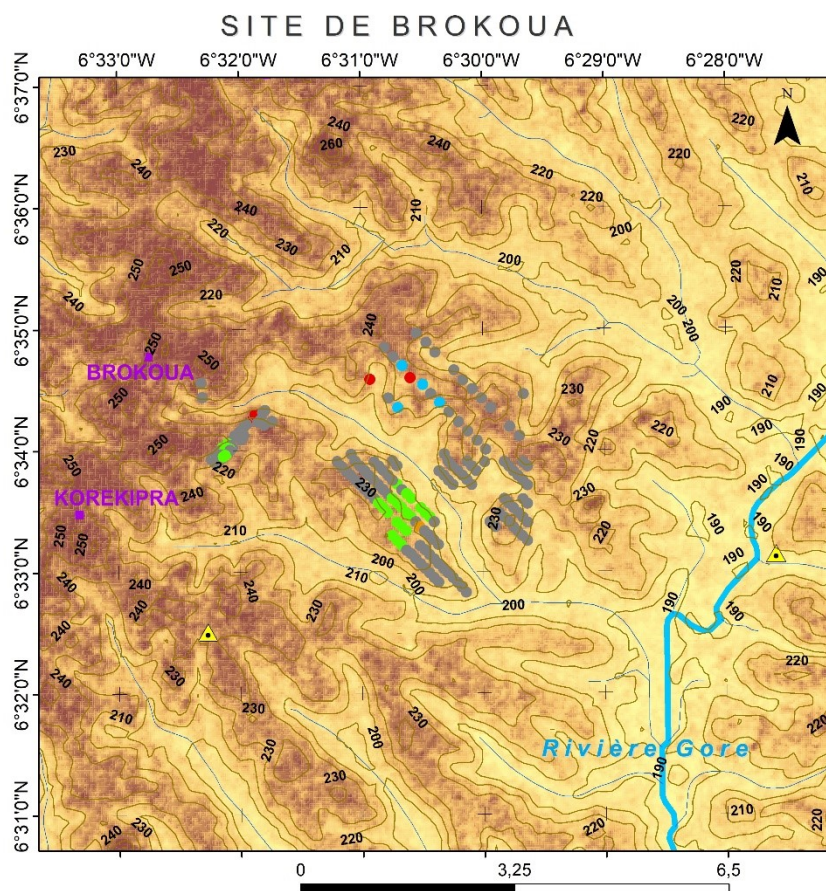


Figure 32: Cartes du relief du site de Bémadi



Légende

▲ Coltan (CT)

■ Localités

Autocorrélation spatiale des échantillons

- Non significative (AL) : Occurences dispersées
- Grappe significative (HH)
- Grappe-point abérant (HL)
- Grappe-point abérant (LH)
- Grappe significative (LL)

— Topographie

Hydrographie

— Marigots

— Rivière

Facteur morphologique (élévation)

Altitude (m)

Élevé : 391

Faible : 19

Figure 33: Cartes du relief des sites de Brokoua et Makua

L'interprétation des cartes des altitudes situe toutes les cibles minérales au niveau des altitudes comprises entre 200 et 393 m (surface de couleur violacée). Ces zones correspondent aux surfaces alluvionnaires, colluvionnaires et éluvionnaires où les dépôts connus ont été identifiés par plusieurs travaux de prospection minière. Les pixels verts à marron représentent les plus bas niveaux topographiques (lits de cours d'eau et ravins). Les pixels de couleur marron à blanc représentent les zones de plus en plus élevées (zones de basfonds, vallées, pénéplaines, flancs de collines, plateaux et crêtes).

Paramètres hydrologiques : Les figures 21 et 22 cartographient la capacité des eaux de ruissellement à mobiliser, transporter, éroder et déposer des sédiments au niveau des sous-bassins versants étudiés. La figure 21 illustre la variation du paramètre d'accumulation de flux.

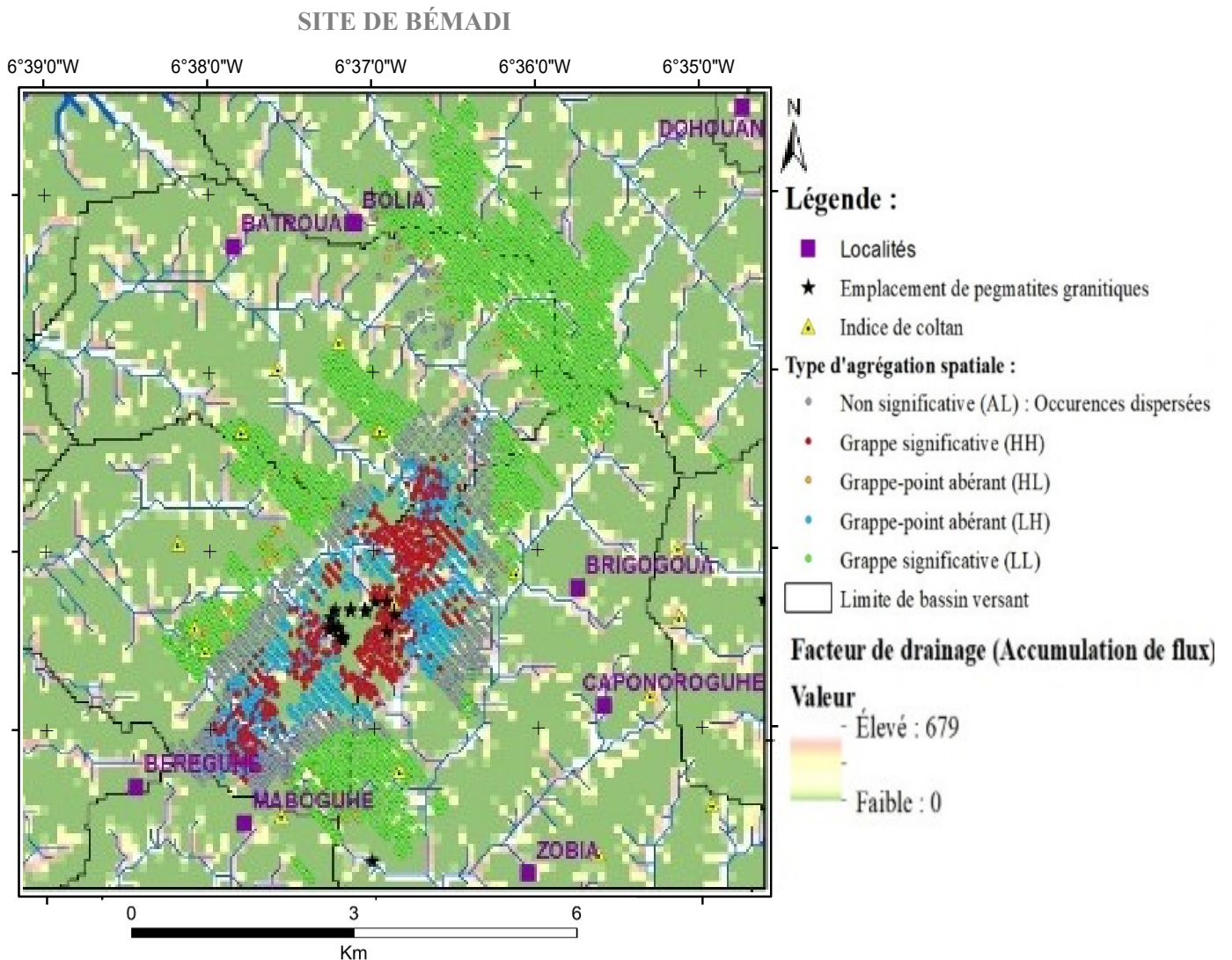
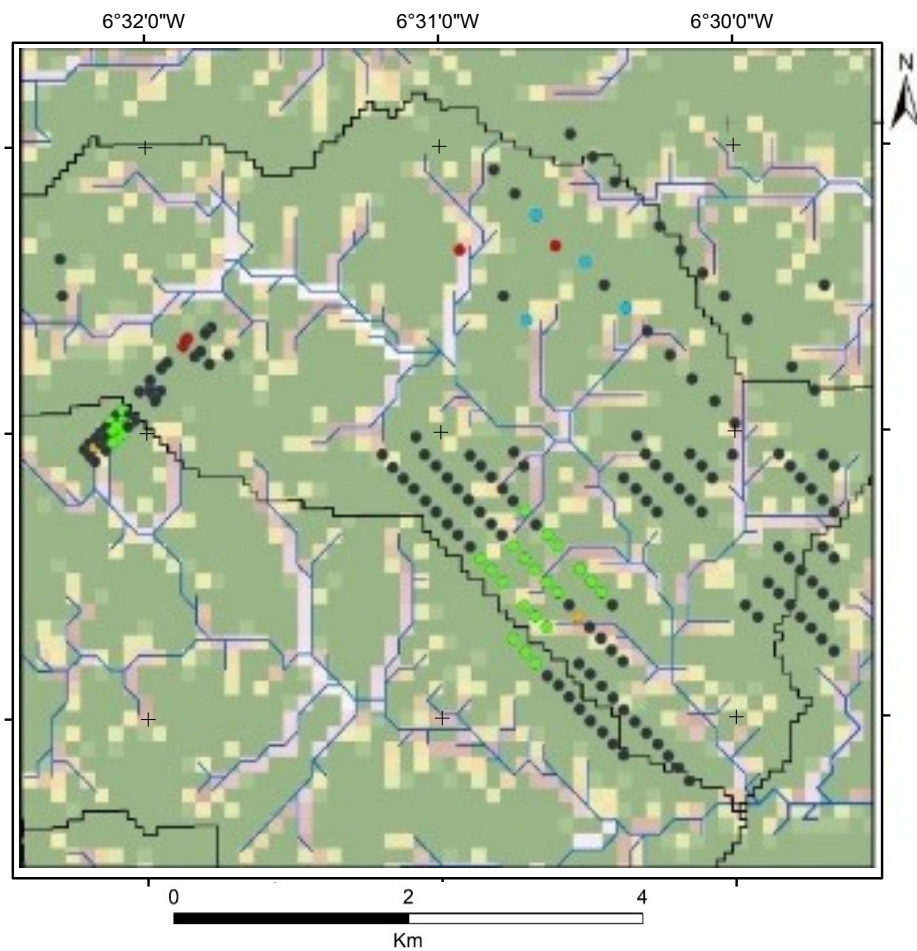
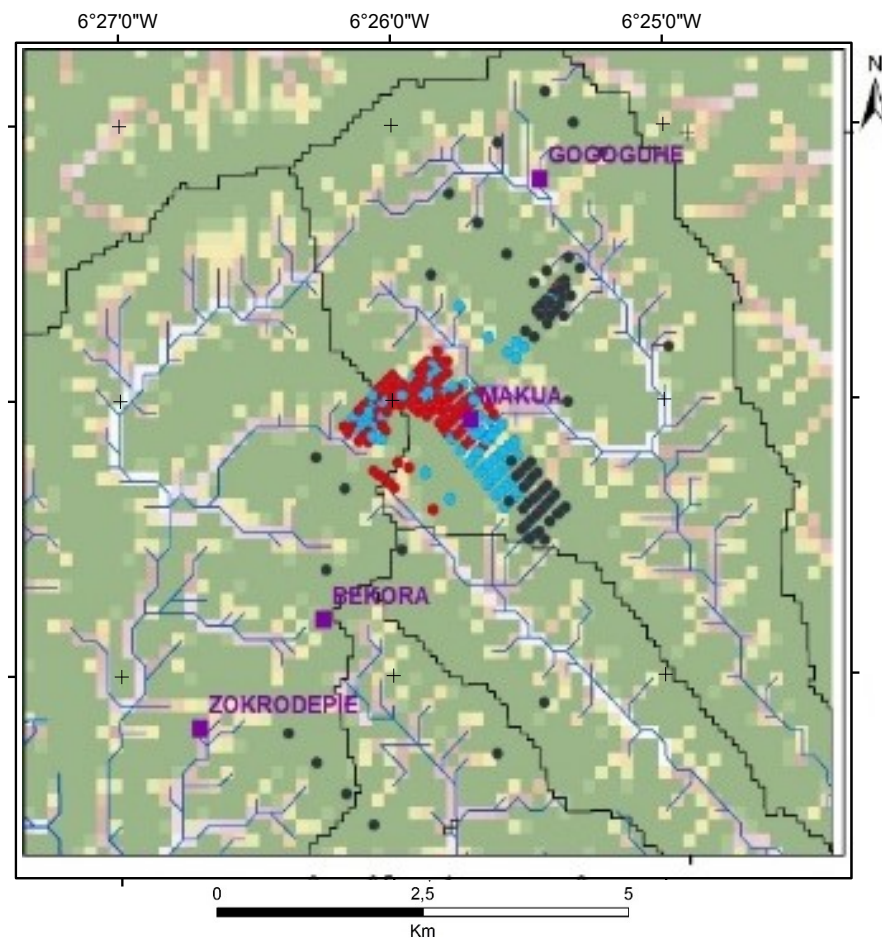


Figure 34: Cartes des zones d'accumulation de flux du site de Bémadi

SITE DE BROKOUA



SITE DE MAKUA



Légende :

- Localités
- Limite de bassin versant

Type d'agrégation spatiale :

- Non significative (AL) : Occurences dispersées
- Grappe significative (HH)
- Grappe-point abérant (HL)
- Grappe-point abérant (LH)
- Grappe significative (LL)

Facteur de drainage (Accumulation de flux)

- Valeur**
- Élevé : 679
 - Faible : 0

Figure 35: Cartes des zones d'accumulation de flux des sites de Brokoua et Makua

L'observation et l'analyse de ces cartes montrent que le potentiel d'accumulation de flux présente des valeurs maximales dans les pixels jaune-orangé à blanc. Ces valeurs correspondent aux lits et berges de marigots ou de rivières responsables de la collecte des eaux de ruissellement des bassins versants. Ces endroits étroits, soumis à un fort processus de drainage, mettent en évidence des zones peu favorables au dépôt de sédiments. En revanche, ce potentiel tend à diminuer au fur et à mesure que l'on s'en éloigne dans la gamme des couleurs vertes. Ces zones très vastes et occupent la grande majorité des sous-bassins versants. Elles sont caractérisées par une bonne capacité de mobilisation et de transport de sédiments. Elles sont favorables au dépôt de sédiments sur de longues distances. La figure 22 illustre la variation du paramètre d'amplitude de flux.

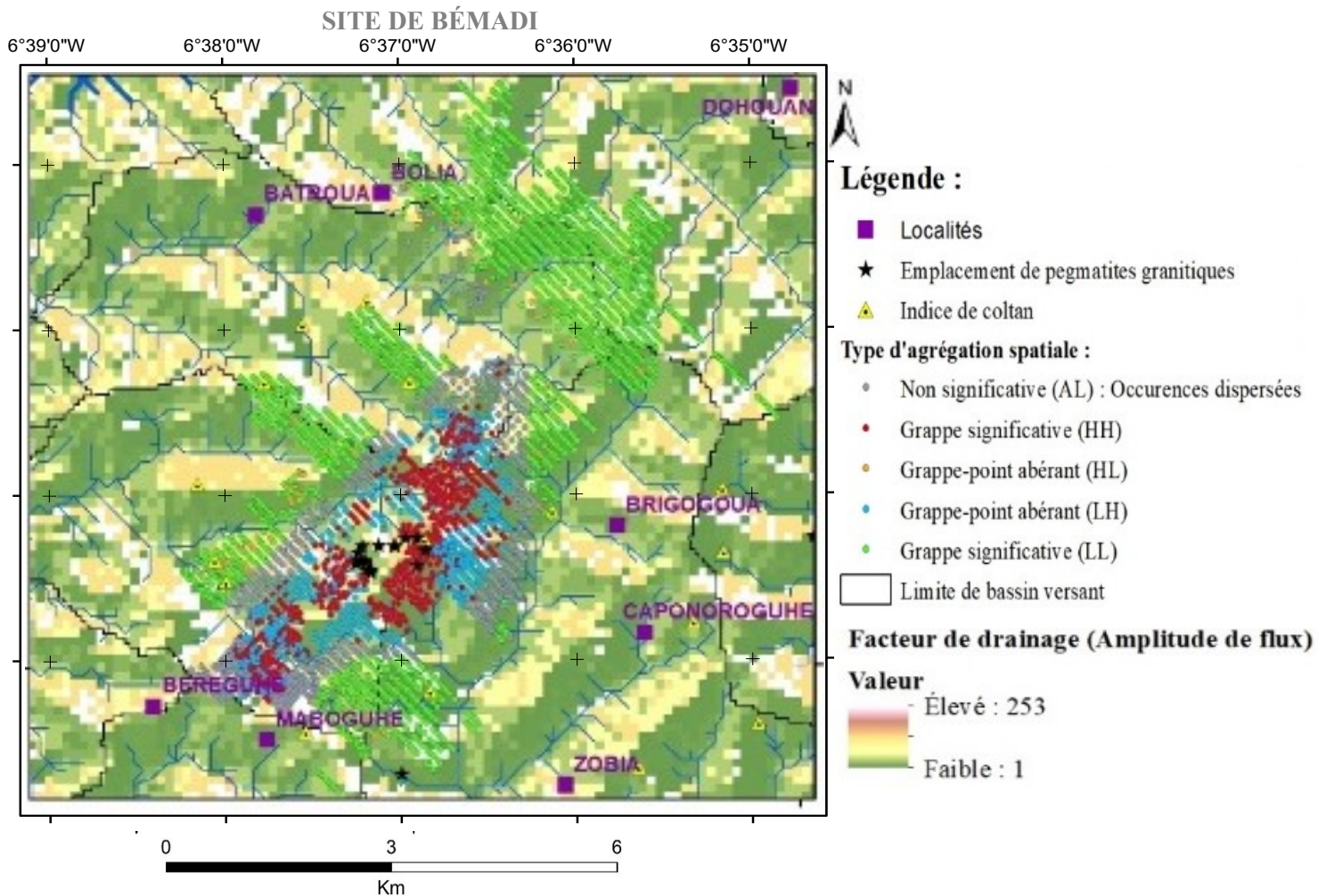
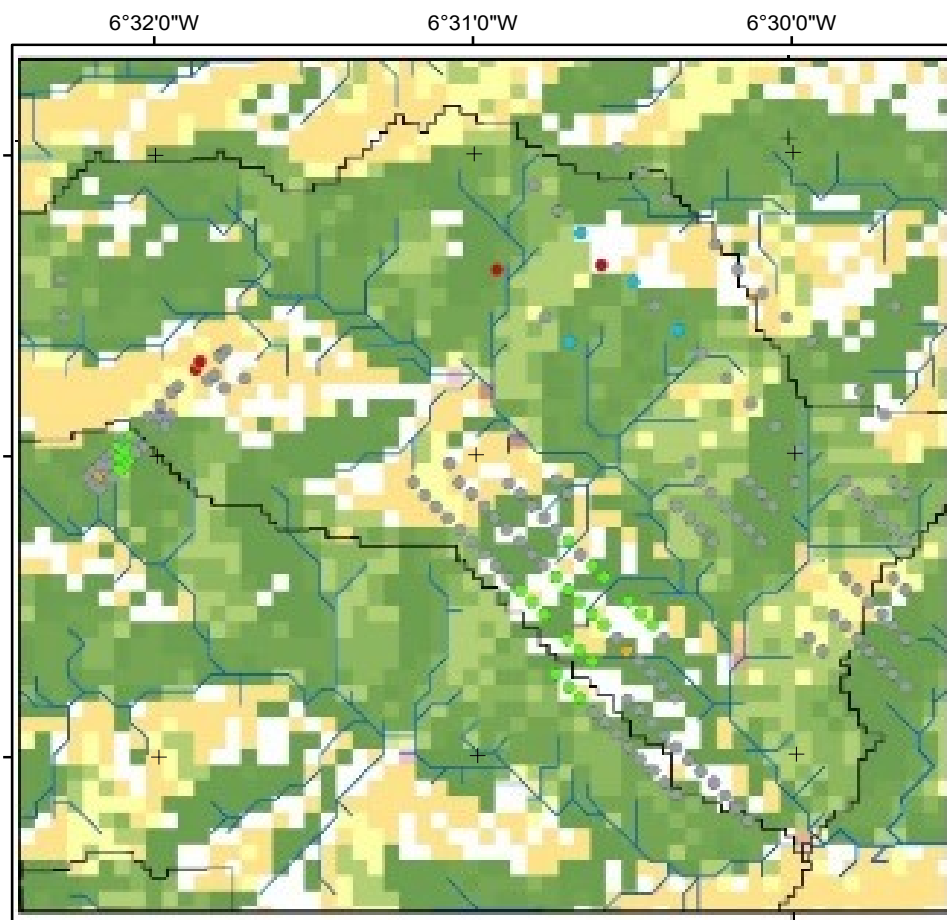


Figure 36: Cartes d'amplitudes et de directions des flux du site de Bémadi.

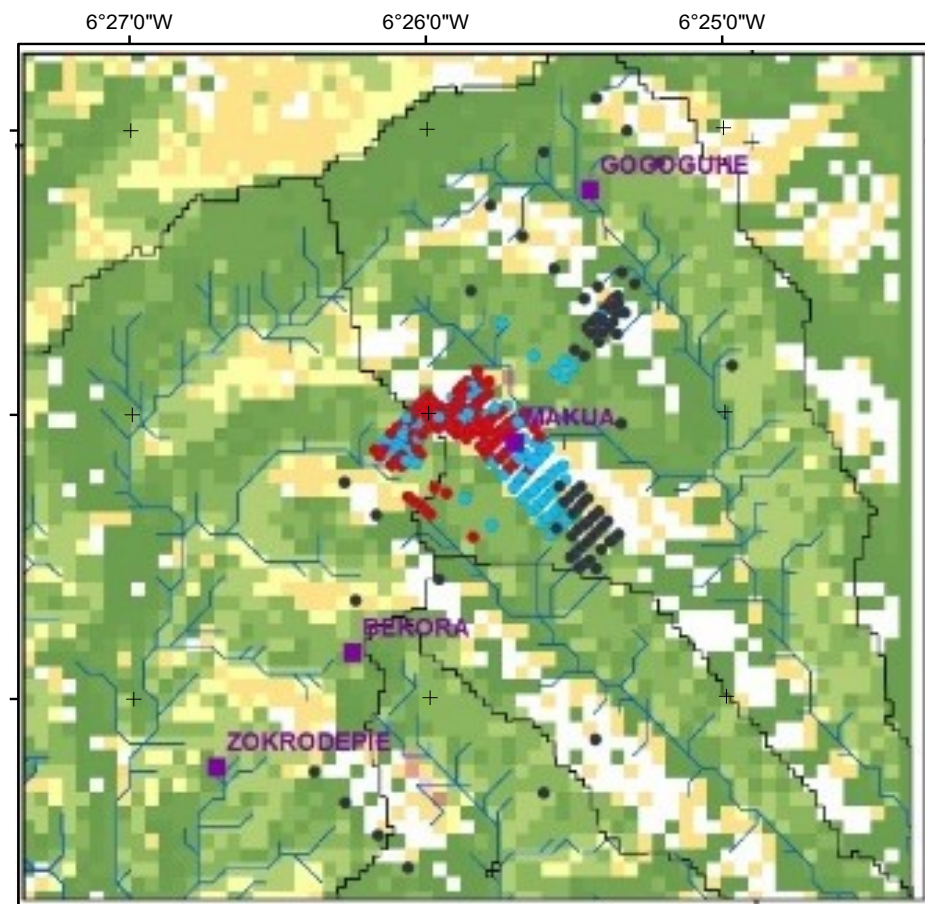


Légende :

- Localités
- Limite de bassin versant

Type d'agrégation spatiale :

- Non significative (AL) : Occurences dispersées
- Grappe significative (HH)
- Grappe-point abérant (HL)
- Grappe-point abérant (LH)
- Grappe significative (LL)



Facteur de drainage (Amplitude de flux)

Valeur

- Élevé : 253
- Faible : 1

Figure 37: Cartes d'amplitudes et de directions des flux des sites de Brokoua et Makua

L'observation et l'analyse des cartes de direction ou d'amplitude d'écoulement mettent en évidence, de part et d'autre des marigots et rivières, des zones d'interfluves (en vert) et d'irrégularités topographiques (crêtes, flancs ou sommets) en jaune – orangé. On observe une nette dissymétrie de la répartition de ces deux classes de couleurs, le long des bordures des cours d'eau. Les zones de couleurs jaune-orangé correspondent aux surfaces élevées et constituent le point de départ des eaux de ruissellement. Les cellules en vert occupent des niveaux topographiques de plus en plus bas. De ce fait, l'amplitude de flux sur ces surfaces atteint des valeurs élevées et diminue peu à peu dans les zones en vert. Les eaux de ruissellement s'écoulent alors des pixels jaunes – orangés, en direction des cellules en vert sur la carte. Enfin, on peut observer que la superficie occupée par les zones jaune-orangé est moins étalée que celle occupée par les zones en vert.

L'ensemble de ces critères spatiaux de preuves sous-jacentes, susceptibles de contrôler ou d'influencer les processus de formation des cibles minérales connues, a été utilisé comme prédicteurs spatiaux. La combinaison des valeurs de ces critères de preuves et leur attribution, en tous points du jeu de données de prospection minérale, a permis de construire notre jeu de données d'apprentissage. Lequel, a été utilisé dans le cadre de l'application des techniques de modélisation prédictive à l'aide des algorithmes de classification automatique.

La prochaine section présente les résultats de cette étape de modélisation prédictive, réalisée sur le jeu de données d'apprentissage, à l'aide des algorithmes de classification automatique (KNN, SVM et DT). Les résultats de précisions et de performances (en termes d'AUC) de tous les modèles entraînés sont d'abord présentés puis analysés. Par la suite, les résultats des différentes analyses comparatives croisées des meilleurs modèles de chaque catégorie sont présentés et analysés, en termes de précision, de valeurs d'AUC et de matrices de confusions. Pour finir, les résultats des tests de validation et de robustesse sur un jeu de données externes sont présentés puis analysés.

4.3 Analyse des résultats d'apprentissage automatique

Les tableaux 4 à 6 présentent les valeurs de précision globale de chaque modèle entraîné et les valeurs AUC estimées pour chaque classe de cible minérale. Il s'agit des résultats obtenus, à la suite de l'application des différentes combinaisons de paramètres d'apprentissage (mentionnés au tableau 2). Le tableau 4 présente les résultats obtenus avec les algorithmes SVM.

Tableau 4 : Résultats des classifications automatiques avec les SVM

Nom du modèle	Méthode utilisée	Valeur de noyau utilisée	Précision globale (%)	Valeur AUC (%)			
				Classe1 (AL)	Classe2 (LL)	Classe3 (LH)	Classe4 (HH/HL)
SVM linéaire	<i>One vs One</i>		41,6	63	58	51	55
	<i>One vs All</i>		39,8	58	63	58	54
SVM à noyau quadratique	<i>One vs One</i>	1	45,2	67	66	62	66
	<i>One vs All</i>	1	42	55	67	63	60
SVM à noyau cubique	<i>One vs One</i>	1	52	71	74	70	71
	<i>One vs All</i>	1	41,8	65	64	63	66
SVM à noyau gaussien fin	<i>One vs One</i>	0,61	61,9	79	85	75	78
	<i>One vs All</i>	0,61	62,4	79	85	75	78
SVM à noyau gaussien moyen	<i>One vs One</i>	2,4	50,5	70	73	65	69
	<i>One vs All</i>	2,4	49,9	63	74	68	68
SVM à noyau gaussien grossier	<i>One vs One</i>	9,8	41,7	65	65	59	58
	<i>One vs All</i>	9,8	42,7	59	65	59	56

L'utilisation des algorithmes SVM sur le jeu de données d'apprentissage, suivant la méthode *One vs One*, montre un classement hiérarchique des pourcentages de précision globale de classification.

Les modèles SVM linéaires et SVM à noyau gaussien grossier sont les moins précis à 42 %, le modèle SVM à noyau quadratique est précis à 45 %, le modèle SVM à noyau gaussien moyen est précis à 51 %, le modèle SVM à noyau cubique est précis à 52 % et le modèle SVM à noyau gaussien fin est le plus précis à 62 %.

L'application de la méthode *One vs All* réduit substantiellement les précisions des classifieurs linéaire, cubique, gaussien moyen et quadratique.

En revanche, la précision augmente légèrement pour le modèle gaussien grossier et reste invariable à 62 % pour le modèle gaussien fin. Suivant cette méthode, le modèle SVM linéaire est précis à 40 %, le modèle SVM à noyau cubique est précis à 42 %, les modèles SVM à noyaux quadratique et gaussien grossier sont précis à 42 % et le modèle SVM à noyau gaussien moyen est précis à 50 %.

L'interprétation de ce tableau fait ressortir que, dépendamment de la méthode de classification utilisée, la précision des modèles SVM à noyau gaussien fin reste inchangée et demeure la plus précise sur le jeu de données d'entrée. Suivant l'analyse des valeurs d'AUC obtenues pour chacun des modèles SVM entraînés, les valeurs d'AUC des modèles à noyau gaussien fin sont les plus élevées. Le modèle à noyau gaussien fin parvient à prédire avec 75 % de probabilité la classe 3 (LH), à 78 % la classe 4 (HH/HL), à 79 % la classe 1 (AL) et à 85 % la classe 2 (LL).

Par conséquent, le modèle le plus précis et le plus performant de la catégorie des SVM est le modèle SVM à noyau gaussien fin.

Au titre des classifieurs KNN, le tableau 5 présente les résultats des classifications automatiques en termes de précision globale. Les valeurs obtenues présentent également un classement hiérarchique selon le type de classifieur utilisé.

Tableau 5 : Résultats des classifications automatiques avec les KNN

Nom du modèle	Paramètres de distance métrique et de poids utilisés	Nombre de voisins	Précision globale (%)	Valeur AUC (%)			
				Classe 1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)
KNN fin	Euclidienne, égal	1	63,7	73	81	67	73
	Cubique, égal	1	63,1	73	80	66	73
	Cosinus, égal	1	62,9	73	80	67	72
KNN moyen	Euclidienne, égal	10	50,5	71	76	69	72
	Cubique, égal	10	49,7	71	75	69	72
	Cosinus, égal	10	49,1	70	75	69	73
KNN grossier	Euclidienne, égal	100	45,9	66	67	64	64
	Cubique, égal	100	45,8	65	66	64	64
	Cosinus, égal	100	45,9	65	67	63	64
KNN cosinus	Euclidienne, égal	10	50,5	71	76	69	72
	Cosinus, égal	10	49,1	70	75	69	73
	Euclidienne, égal	10	50,5	71	76	69	72
	Cubique, égal	10	49,7	71	75	69	72
	Euclidienne, égal	10	50,5	71	76	69	72
KNN pondéré-euclidienne (IDW)	Euclidienne, Inverse de la distance	10	65,3	82	89	75	78
KNN pondéré-euclidienne (IDW2)	Euclidienne, Inverse de la distance au carré	10	65,3	82	89	75	78

Quelle que soit la méthode de pondération utilisée, la précision des trois modèles KNN grossier reste constante à 46 %. Ces modèles grossiers utilisent les 100 voisins les plus proches et sont les moins précis de cette catégorie. Les modèles KNN moyen et KNN cosinus entraînés avec un paramètre de distance métrique cosinus et une pondération métrique égale, sont tous deux précis à 49 %. Les modèles KNN cubique et KNN moyen entraînés avec un paramètre de distance métrique cubique et une pondération métrique égale, sont tous deux précis à 50 %. Les modèles KNN moyen, KNN cosinus, KNN cubique et KNN pondéré-euclidienne utilisant des paramètres de distance métrique euclidienne, avec une pondération égale, sont tous les trois précis à 51 %.

Il ressort de cela que la variation du paramètre de distance métrique entraîne une amélioration subséquente de la précision de classification, au fur et à mesure que le nombre de voisins les plus proches utilisés pour l'apprentissage diminue. En effet, les modèles KNN fin entraînés avec 1 seul voisin le plus proche et des distances métriques cubique, euclidienne ou cosinus, avec une pondération égale, présentent une nette augmentation de leurs précisions globales, à environ 63 et 64 %.

Cependant, l'utilisation d'un paramètre de pondération, basé sur le calcul de l'inverse de la distance séparant les 10 plus proches voisins, donne de meilleures performances pour notre jeu de données. En effet, une précision de 65 % a été obtenue avec les modèles KNN pondéré-euclidienne (IDW) et KNN pondéré-euclidienne (IDW2).

En termes des valeurs d'AUC, les résultats confirment qu'un modèle KNN pondéré-euclidienne (IDW) ou (IDW2) présente les meilleures performances prédictives pour les classes cibles minérales. Le modèle parvient à prédire 75 % des cibles de la classe 3 (LH), 78 % des cibles de la classe 4 (HH/HL), 82 % des cibles de la classe 1 (AL) et 89 % des cibles de la classe 2 (LL). Par conséquent, les modèles les plus précis et les plus performants de cette catégorie, sont les modèles KNN pondéré-euclidienne (IDW) et (IDW2).

Au titre des classifieurs utilisant les DT, les tableaux 6 et 7 présentent les résultats obtenus en termes de précisions et de performances des classifieurs.

Les résultats obtenus avec des algorithmes d'arbres de décision unique sont présentés au tableau 6.

Tableau 6 : Résultat des classifications automatiques avec les DT uniques

Nom du modèle	Paramètre de critère de division utilisé	Nombre maximum de divisions	Précision globale (%)	Valeur AUC (%)			
				Classe1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)
DT simple	Indice de diversité de Gini	4	45,8	58	60	56	59
	Règle de Twoing	4	45,7	58	60	56	59
	Réduction maximum de déviance	4	45,5	58	60	56	59
DT moyen	Indice de diversité de Gini	20	46,6	62	64	61	61
	Règle de Twoing	20	46,3	62	62	61	61
	Réduction maximum de déviance	20	46,9	62	63	61	61
DT complexe	Indice de diversité de Gini	100	54	76	77	71	73
	Règle de Twoing	100	52,8	73	76	72	70
	Réduction maximum de déviance	100	52	74	76	71	71

L'analyse du tableau 6 montre que, dépendamment du critère de division utilisé, les modèles prédicteurs utilisant des arbres de décision simple (avec un nombre maximum de 4 divisions) et moyen (avec un nombre maximum de 20 divisions) montrent ou sinon très peu, une invariabilité du pourcentage de précision globale (de l'ordre de 46 et 47 %). Ces modèles sont les moins performants de cette catégorie.

Cependant, la précision et la performance des modèles DT complexes sont influencées par l'application des différents critères de division (réduction maximum de déviance, règle de Twoing et indice de diversité de Gini). La précision augmente à 52 % avec la réduction maximale de déviance, puis à 53 % avec la règle de Twoing et à 54 % avec l'indice de diversité de Gini. L'analyse des valeurs d'AUC confirme cette tendance. En effet, le modèle DT complexe parvient à prédire 71 % des cibles minérales de la classe 3 (LH), 73 % des cibles minérales de la classe 4 (HH/HL), 76 % des cibles minérales de la classe 1 (AL) et 77 % des cibles minérales de la classe 2 (LL). Par conséquent, le modèle le plus

précis et le plus performant de la catégorie des DT uniques est le modèle DT complexe utilisant l'indice de Gini comme critère de division. De plus, les performances des DT uniques augmentent avec le nombre de divisions maximum et la complexité de l'arbre.

Un autre type d'algorithmes d'arbres de décision a été étudié. Il s'agit des arbres de décisions assemblés selon une structure beaucoup plus complexe. Le tableau 7 en présente les résultats obtenus.

Tableau 7: Résultats des classifications automatiques avec les DT assemblés

Nom du modèle	Nombre maximum de divisions et nombre d'apprenants	Taux d'apprentissage	Précision globale (%)	Valeur AUC (%)			
				Classe1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)
Ensemble DT renforcés	20 / 30	0,1	47,4	68	69	66	65
	20 / 30	1	50,3	76	76	73	71
	10 / 60	0,1	46,9	67	69	64	63
	20 / 30	0,1	47,4	68	69	66	65
	20 / 60	1	51,9	77	78	75	73
	20 / 100	0,1	48	71	73	70	68
	20 / 100	1	54	76	79	75	73
	10 / 100	1	49,7	73	75	72	70
	10 / 100	0,1	46,9	68	69	65	64
	20 / 500	1	58,1	75	77	74	71
	20 / 500	0,1	51,5	77	77	74	73
	20 / 800	0,1	53,1	77	78	75	74
	20 / 800	1	58,5	73	75	73	68
Ensemble DT ensachés (Random forest)	10/30	0,1	67,3	87	91	80	81
	10/60	0,1	68,2	87	91	81	80
	10/500	0,1	68,4	88	91	82	81
	30/500	0,1	68,8	88	91	81	81
	30/100	0,1	69	87	91	82	81
	10/100	0,1	68,3	87	91	81	81
	20/30	0,1	67,2	86	90	80	80
	20/60	0,1	67,5	87	91	80	81
	20/100	0,1	68,3	87	91	81	81
	20/500	0,1	68,4	88	91	81	81
	30/500	0,1	68,5	87	91	81	81
	30/60	0,1	68,1	87	91	81	81
	30/30	0,1	67,3	86	90	80	80

La technique de combinaison de plusieurs paramètres, notamment le nombre maximum de divisions, le nombre d'apprenants et le taux d'apprentissage ont été appliqués, quand ce fut possible, sur le jeu de données d'entraînement. Cela a permis d'évaluer l'influence de la variation et de l'association de ces paramètres sur la performance des différents classifieurs. L'analyse du tableau 7 montre de prime abord que les modèles DT renforcés sont moins précis que les modèles *Random Forest*. Le modèle prédicteur DT renforcés le plus précis obtient un score de 59 %, quand le modèle RF le plus précis obtient un score de 69 %.

Au niveau des DT renforcés, l'augmentation du nombre maximum de divisions de 10 à 20 améliore les précisions de classification de 47 % à 53 %, pour un taux d'apprentissage fixé à 0,1. La même tendance est observée pour le taux d'apprentissage fixé à 1, où la précision augmente de 50 à 59 %. Il faut également noter une certaine influence de la variation du nombre d'apprenants (30, 60, 100, 500 et 800) dans ce classement. En effet, il apparaît pour un taux d'apprentissage et un nombre de divisions constant que la précision tend à croître avec le nombre d'apprenants. Car, au taux d'apprentissage de 0,1, le modèle DT renforcés utilisant 20 divisions maximum et 800 apprenants obtient le meilleur score (53 %) de sa catégorie. Il en est de même au taux d'apprentissage de 1, où le modèle DT renforcés utilisant 20 divisions maximum et 800 apprenants obtient un score de 59 % contre 50 % pour le modèle DT renforcés avec 20 divisions et 30 apprenants.

Au niveau des modèles RF, pour des raisons propres au logiciel utilisé dans cette étude, seul un taux d'apprentissage de 0,1 a pu être testé lors de la formation des algorithmes. Contrairement aux DT renforcés, la combinaison et la variation des différents paramètres utilisés n'ont eu que très peu d'influence sur les précisions et les performances des modèles entraînés. En effet, un modèle entraîné avec 10 divisions maximales et 100 apprenants est tout aussi précis et performant à 69 % qu'un modèle entraîné avec 30 divisions et 500 apprenants. Les scores de tous les algorithmes RF testés varient entre 67 et 69 % et sont nettement supérieurs à tous les résultats obtenus avec les DT renforcés. Dans cette catégorie, les modèles DT ensachés avec 30 divisions et 500 apprenants et le modèle DT ensachés avec 30 divisions et 100 apprenants ont performé respectivement et exactement à 68,8 et 69,0 %.

Par conséquent, le meilleur modèle a été sélectionné sur la valeur décimale du pourcentage de précision globale et suivant les valeurs d'AUC prédites par ces deux modèles concurrents. Le modèle RF (30/100) est capable de prédire 87 % des cibles minérales de la classe 1 (AL) contre 88 % pour le modèle RF (30/500). Les deux modèles prédisent respectivement 91 % et 81 % des cibles minérales de la classe 2 (LL) et 4 (HH/HL). Le modèle RF (30/100) prédit 82 % des cibles minérales de la classe 3 (LH) contre 81 % pour le modèle concurrent. Compte tenu de l'importance accordée dans cette étude aux classes 2, 3 et 4 et sur la base de ces critères, c'est le modèle RF (30/100) qui obtient les meilleures valeurs d'AUC. Par conséquent, le modèle le plus précis et le plus performant de la catégorie des DT assemblés, est le modèle prédicteur *Random Forest* construit avec 30 divisions maximales et 100 apprenants.

4.4 Évaluations des meilleurs modèles prédicteurs

Les trois sections suivantes présentent les résultats des trois approches comparatives croisées, utilisées pour évaluer les meilleurs modèles prédicteurs de chaque catégorie, en termes de précision globale, de valeurs d'AUC, de VPP, de VPN, de TVP, de TVN et de nombres d'observations bien et mal classées.

4.4.1 Analyse des modèles en termes de précisions globales

Les résultats des précisions globales de classification ainsi que les valeurs d'AUC obtenues pour chaque meilleur modèle des trois catégories sont résumés au tableau 8.

Tableau 8: Précisions et valeurs d'AUC des trois meilleurs modèles prédicteurs.

Nom du modèle	Précision globale (%)	Valeur AUC (%)			
		Classe1 (AL)	Classe2 (LL)	Classe3 (LH)	Classe4 (HH/HL)
SVM à noyau gaussien fin	62,4	79	85	75	78
KNN pondéré-euclidienne (IDW)	65,3	82	89	75	78
Random Forest	69	87	91	82	81

L'analyse comparative de ces valeurs montre que le meilleur modèle prédicteur SVM à noyau gaussien fin est précis à 62,4 %. Le meilleur modèle prédicteur KNN (pondéré-euclidienne (IDW)) est précis à 65,3 % tandis que le meilleur modèle prédicteur *Random Forest* est précis à 69 %.

De cette première analyse comparative, en termes de précision globale de classification, il ressort de cela que le modèle *Random Forest* est plus précis sur les données d'apprentissage que le meilleur modèle KNN.

À son tour, le meilleur modèle KNN est plus précis sur les données d'apprentissage que le meilleur modèle SVM. Dans l'optique d'approfondir cette analyse comparative, les valeurs d'AUC et les profils des courbes ROC de chacun des modèles ont été analysés et interprétés.

4.4.2 Analyse des modèles en termes d'AUC

L'analyse des valeurs d'AUC offre plusieurs avantages. Le premier est que l'AUC est invariante d'échelle. Le second est que celle-ci mesure la qualité du classement des prédictions, plutôt que leurs valeurs absolues.

Finalement, le troisième est que l'AUC mesure la qualité des précisions du modèle, quel que soit le seuil de classification sélectionné.

Les figures 38, 39 et 40 présentent respectivement les courbes ROC et les valeurs AUC obtenues respectivement pour le meilleur modèle RF, KNN et SVM.

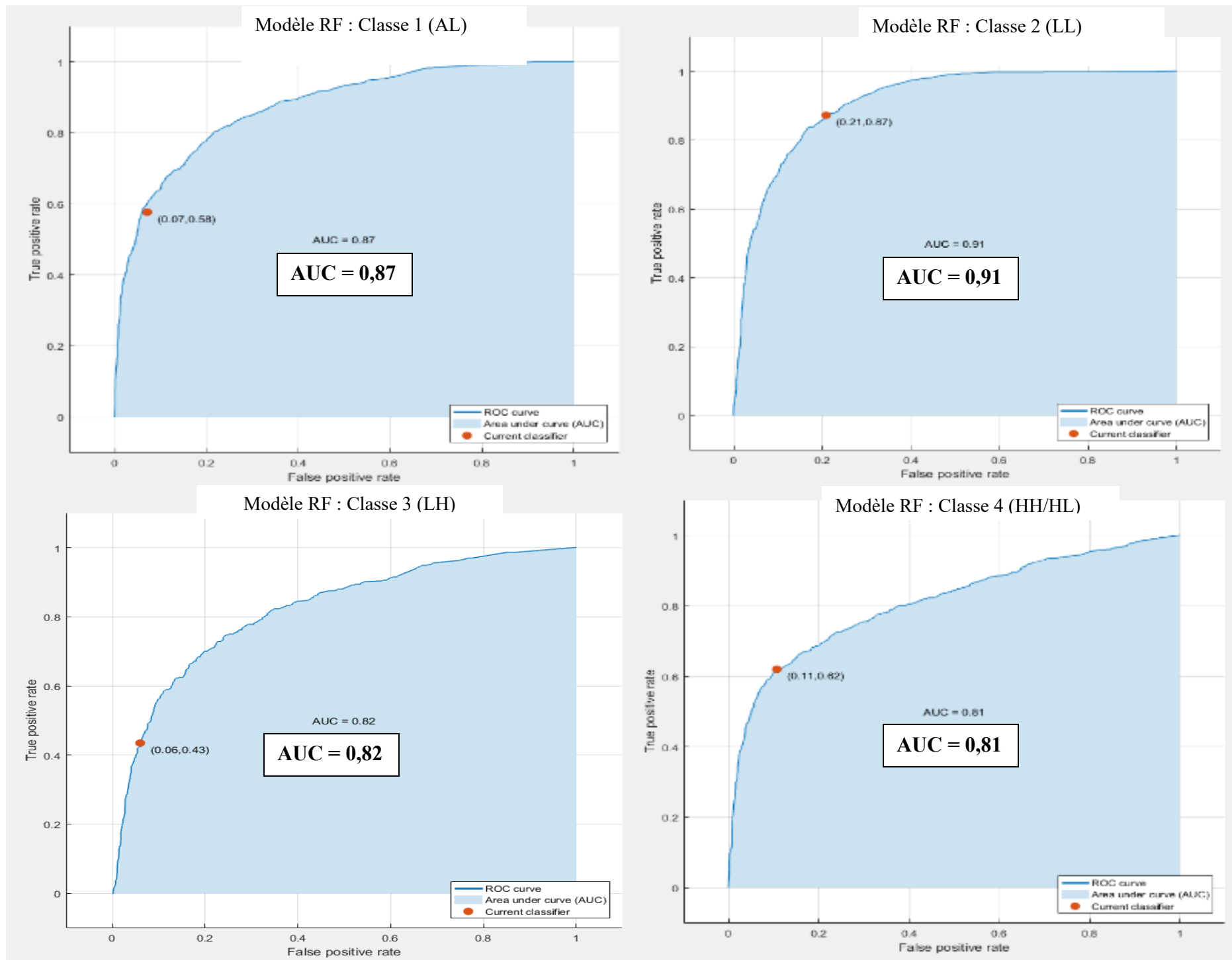


Figure 38: Courbes ROC et valeurs AUC du meilleur modèle ranfom forest (RF)

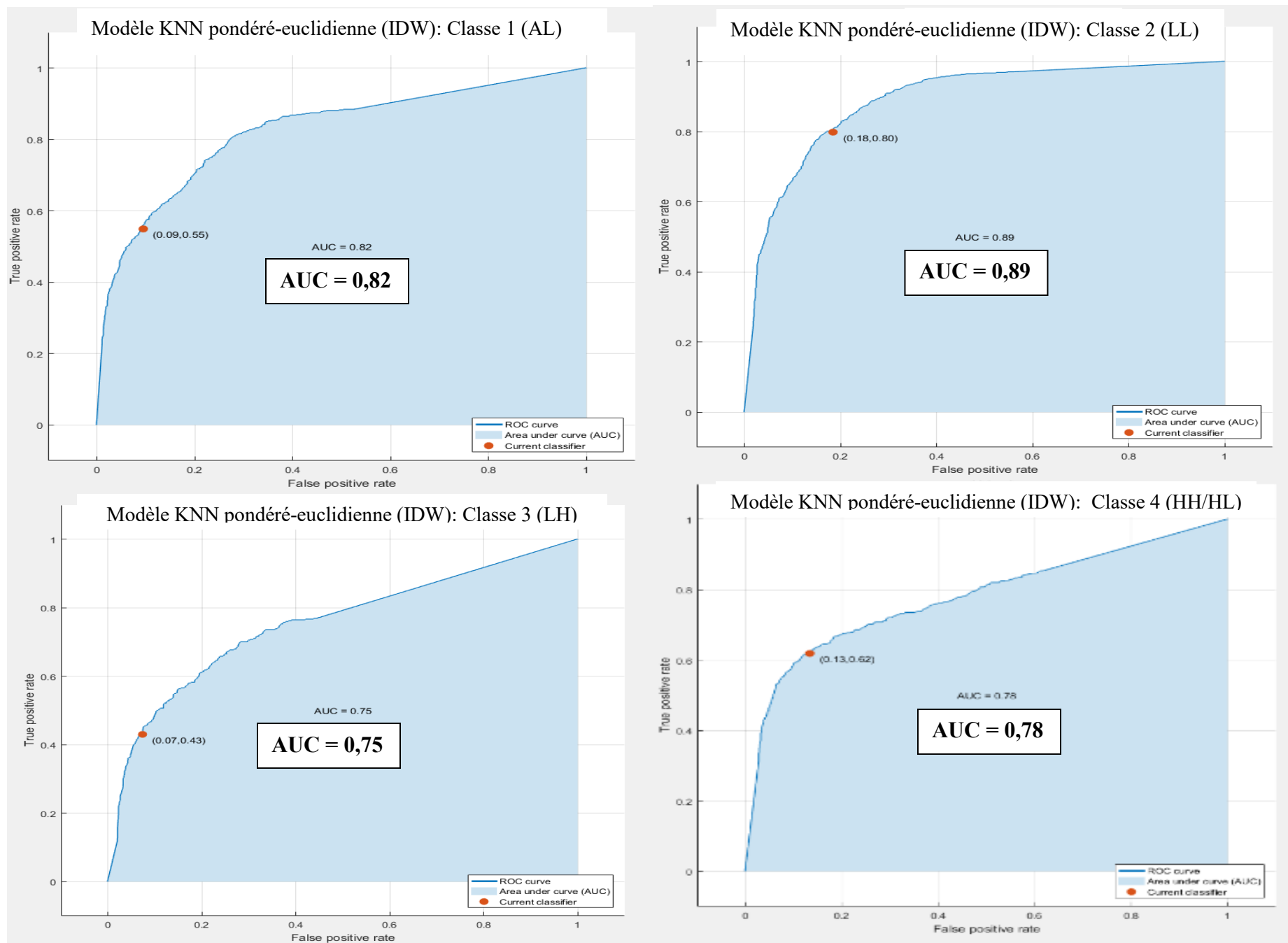


Figure 39 : Courbes ROC et valeurs AUC du meilleur modèle KNN (pondéré-euclidienne (IDW))

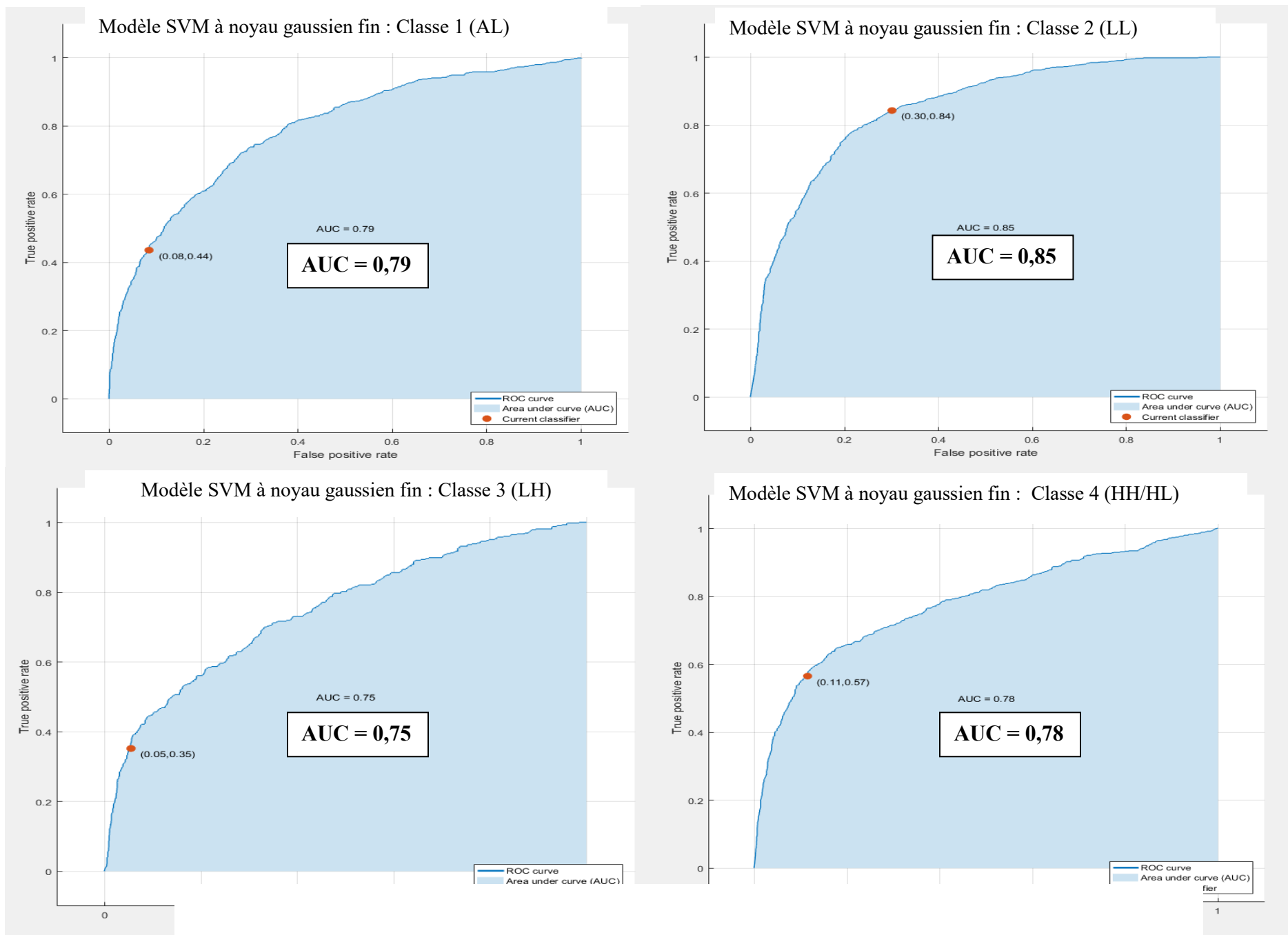


Figure 40: Courbes ROC et valeurs AUC du meilleur modèle SVM à noyau gaussien fin

La figure 38 montre que le meilleur modèle prédicteur *Random Forest* est capable de prédire 81 % des cibles minérales de type (HH/HL), 82 % des cibles minérales de type (LH), 87 % des cibles minérales de type (AL) et 91 % des cibles minérales de type (LL). L'allure convexe des quatre courbes ROC obtenues et les valeurs d'AUC comprises entre 82 et 91 % confirment que le modèle RF donne de bonnes probabilités pour classer toutes les cibles minérales à des taux satisfaisants. Cependant, le modèle est plus spécifique et plus sensible à 91 % pour la classe 2 (LL).

La figure 39 montre que le meilleur modèle prédicteur KNN pondéré-euclidienne (IDW) est capable de prédire 75 % des cibles minérales de type (LH), 78 % des cibles minérales de type (HH/HL), 82 % des cibles minérales de type (AL) et 89 % des cibles minérales de type (LL). L'allure convexe des quatre courbes ROC obtenues et les valeurs d'AUC comprises entre 75 et 89 % confirment que le modèle KNN donne également de bonnes probabilités pour classer les différents types de cibles minérales. Cependant, le modèle est plus spécifique et plus sensible à 89 % pour la classe 2 (LL).

La figure 40 montre que le meilleur modèle prédicteur SVM à noyau gaussien fin est capable de prédire 75 % des cibles minérales de type (LH), 78 % des cibles minérales de type (HH/HL), 79 % des cibles minérales de type (AL) et 85 % des cibles minérales de type (LL). L'allure convexe des quatre courbes ROC obtenues et les valeurs d'AUC comprises entre 75 et 85 % confirment que le modèle SVM donne également, de bonnes probabilités pour classer les quatre types de cibles minérales. Le modèle est plus spécifique et plus sensible à 85 % pour la classe 2 (LL).

De façon comparative, les valeurs d'AUC obtenues avec le RF sont supérieures à celles du KNN, lesquelles sont supérieures à celles estimées par le SVM. En résumé, l'analyse comparative des courbes ROC et des valeurs d'AUC confirme que les trois meilleurs modèles prédicteurs offrent de bonnes performances de classification notamment pour la classe 2 (LL). Cependant, comparativement à la classe 2, il faut noter une réduction sensible de la capacité de ces modèles à classer et prédire les cibles minérales des classes (LH); (HH/HL) et (AL).

Pour tenter d'apporter un éclairage sur ce comportement commun aux trois modèles, un histogramme comparatif des valeurs AUC estimées par chaque modèle a été construit. La figure 26 présente l'histogramme comparatif des performances de chaque modèle, en termes de valeurs d'AUC estimées pour les différents types de cibles minérales. Dans le but de faciliter la compréhension de cette analyse, un code de couleur a été adopté pour chaque catégorie d'algorithmes. La couleur bleue est assignée aux résultats du meilleur modèle SVM, l'orange à ceux du meilleur modèle KNN et le gris au meilleur modèle RF.

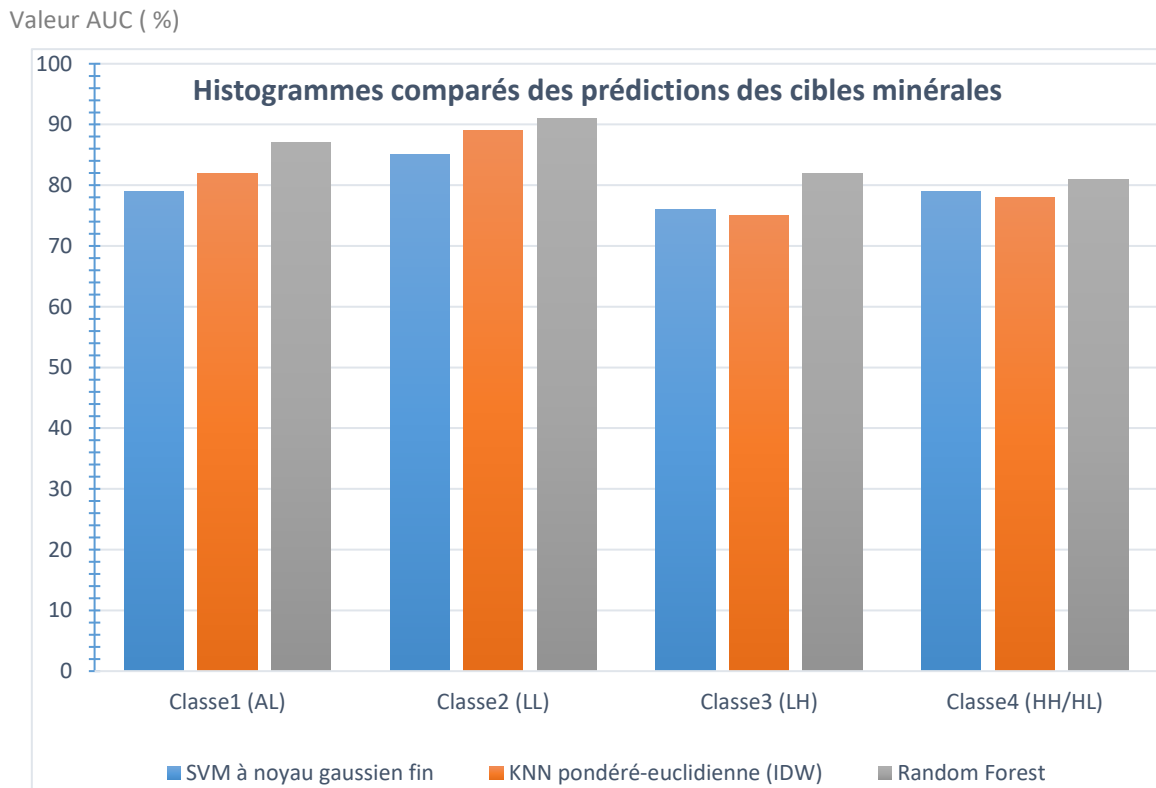


Figure 41: Histogrammes comparés des performances prédictives

Les valeurs en ordonnées du graphique sont les AUC exprimées en pourcentage. L'histogramme en bleu représente les performances du modèle SVM pour la prédiction de chaque type de cible minérale. Les histogrammes, en orange et gris, représentent respectivement les valeurs d'AUC estimées par les modèles KNN et RF.

Au point de vue de la performance prédictive de chaque type de cible minérale, l'observation des hauteurs de chaque barre, montre que la capacité prédictive des trois modèles est la plus élevée pour la classe 2. Mais, cette capacité diminue au fur et à mesure pour la classe 1 et 3.

L'analyse de la composition numéraire de ces classes, dans le jeu de données d'apprentissage (tableau 9) permet d'émettre une hypothèse et d'expliquer un tel déséquilibre au niveau des prédictions de ces modèles prédicteurs.

Tableau 9: Composition numéraire des classes d'apprentissage.

Jeu de données	Classe1		Classe2		Classe3		Classe4		Total
Initial	789	21 %	513	14 %	1542	42 %	894	24 %	3709
Apprentissage	617	23 %	1234	46 %	423	16 %	693	26 %	2697

En effet, l'analyse de la composition numéraire montre que la classe 2 (LL) est composée de 1234 entités soit 46 % de l'effectif du jeu de données d'apprentissage. Tandis que la classe 3 est composée de 423 entités, soit 16 % du jeu de données d'apprentissage. Les classes 1 et 4 sont composées respectivement de 617 entités (soit 23 %) et 693 entités (soit 26 %) du jeu de données d'apprentissage. Un tel déséquilibre numéraire dans les données d'entraînement pourrait engendrer un déséquilibre dans la formation des algorithmes et éventuellement affecter la capacité prédictive des modèles résultants. D'abord, il convient de constater que les trois modèles présentent les valeurs d'AUC les plus élevées pour la classe 2 composée de 46 % du jeu de données d'apprentissage. Ensuite, les valeurs d'AUC du RF pour les classes 1, 3 et 4 sont respectivement de 87, 82 et 81 % (pour une composition numéraire respective de 23, 16 et 26 %), quand celles du KNN sont respectivement de 82, 75 et 78 % et celles du SVM sont respectivement de 79, 75 et 78 %.

Dans cette étude, il ressort à première vue que plus la composition numéraire de la classe augmente, plus la valeur d'AUC estimée augmente. Secundo, plus le modèle est précis, moins il est sensible à la structure numéraire du jeu de données d'apprentissage. Cela

signifie que les modèles SVM et KNN seraient beaucoup plus sensibles à la structure du jeu de données d'apprentissage que le modèle prédicteur RF.

D'un point de vue statistique, il est intéressant d'observer et de constater, à l'analyse de l'histogramme de la figure 41, une différence peu significative (de l'ordre de $\pm 10\%$) des performances prédictives des trois (03) meilleurs modèles identifiés. En effet, tous les algorithmes utilisés présentent des prédictions supérieures à 70 % pour toutes les classes présentées. Cela est déjà la preuve que la démarche proposée semble être robuste. De ce fait, les algorithmes RF, SVM et KNN retenus sont tous susceptibles de donner des résultats avec des taux satisfaisants et supérieurs à 70 %. Mais, l'objectif de ce travail est de trouver l'algorithme le plus optimal possible pour modéliser la distribution géostatistique des classes de cibles minérales proposées dans les jeux de données. Ce qui permettra de prédire avec les probabilités les plus élevées possible, de nouvelles zones potentiellement prospectives et de générer *in fine* des cartes de potentialités minérales, les plus précises. Par conséquent, c'est le modèle RF (avec des taux de prédiction de classes, supérieurs ou égaux à 80 %, qui semble (à ce stade de l'étude) être préférable pour prédire correctement la répartition spatiale d'un maximum d'emplacements de cibles minérales.

D'autre part, il importe de tenir compte dans l'évaluation des performances d'un algorithme, des résultats réels de classification des entités du jeu de données d'apprentissage. Car, les valeurs d'AUC et les précisions globales obtenues sont des estimations de la performance de ces modèles. Tandis que les matrices de confusions résultantes des classifications dressent un état (en termes de nombre d'observations bien ou mal classées) de la capacité réelle d'un modèle à classer et prédire l'appartenance d'une entité à une classe. Ce qui permet de jauger la pertinence réelle des classifieurs utilisés. En outre, il faut noter que l'indice de précision globale et le kappa ne tiennent pas compte de la distance entre les classes. En effet, le kappa évalue essentiellement la qualité de la classification effectuée, par rapport à l'attribution aléatoire de valeurs. Il détermine si la classification a été meilleure que la statistique aléatoire. Et la précision globale d'une classification est un indice optimiste des performances du classifieur.

Dans cette optique, une troisième étape d'analyse comparative a été réalisée, pour permettre d'évaluer la capacité prédictive réelle de chaque meilleur modèle (KNN, SVM

et RF) sur les entités composant les quatre types de cibles minérales. La section suivante présente les résultats de calcul des F-scores utilisés pour obtenir une estimation réelle de la précision de classification, en cas de dissymétrie dans les données d'entraînement.

4.4.3 Analyse des modèles en termes de matrices de confusions

Chaque classification automatique produit trois matrices de confusions. La première présente le nombre d'observations bien classées et mal classées. Les lignes correspondent à la classe prédite (classe de sortie) et les colonnes correspondent à la vraie classe (classe cible). Les cellules diagonales correspondent à des observations correctement classées. Les cellules non diagonales correspondent à des observations mal classées. Le nombre d'observations ou le pourcentage du nombre total d'observations sont indiqués dans chaque cellule.

La seconde et la troisième matrice présentent les valeurs des taux de vrais positifs (TVP) et de faux négatifs (TVN) calculées pour chaque classe. Les deux lignes au bas des tableaux montrent les pourcentages de tous les exemples censés appartenir à chaque classe, qui sont correctement (ligne du dessus) et incorrectement classés (ligne en dessous). Ces métriques sont respectivement, la précision (VPP) et le taux de découvertes erronées (VPN). Les colonnes à l'extrême droite des tableaux montrent les pourcentages de tous les exemples appartenant à chaque classe, qui sont correctement (à savoir, le rappel TVP) et incorrectement classés (à savoir, le taux de faux négatif TVN).

Les matrices de confusions comparées de chacun de ces modèles prédictifs, sont présentées aux tableaux 10, 11 et 12. Dans le but de faciliter la compréhension de ces analyses, le même code de couleur a été conservé pour chaque catégorie d'algorithme. La couleur bleue est assignée aux résultats du meilleur modèle SVM, l'orange à ceux du meilleur modèle KNN et la grise au meilleur modèle RF.

Tableau 10: Matrices de confusions comparées en termes de nombre d'observation

Vraie classe	Classe prédite SVM				Classe prédite KNN				Classe prédite RF			
	1	2	3	4	1	2	3	4	1	2	3	4
1	269	248	43	57	339	155	52	71	356	169	38	54
2	88	1041	15	90	109	986	24	115	69	1077	15	73
3	61	100	149	113	61	63	182	117	49	75	184	115
4	48	173	80	392	52	101	111	429	47	117	99	430

L'observation de ces trois matrices, révèle un ordre de bon classement selon les différentes classes de cibles minérales. En premier vient la mieux classée par les trois modèles, à savoir la classe 2 de type (LL), avec 1077 observations pour le modèle RF, 1041 pour le modèle SVM et 986 pour le modèle KNN. La seconde est la classe 4 (HH/HL) avec 430 observations pour le RF, 429 pour le KNN et 392 pour le SVM. La troisième est la classe 1 (AL) avec 356 observations pour le RF, 339 pour le KNN et 269 pour le SVM. Et, la dernière est la classe 3 (LH), avec 184 observations pour le RF, 182 pour le KNN et 149 pour le SVM.

Comparativement aux modèles KNN et SVM, le modèle RF obtient pour toutes les classes les plus grands nombres de bons classements réels. À l'exception de la classe 2 (986 contre 1041), le modèle KNN obtient un plus grand nombre de bonnes classifications que le SVM. Une fois de plus, la structure du jeu de données d'apprentissage semble se refléter sur le nombre d'observations bien classées.

Le tableau 11 présente les matrices de confusions comparées en termes de taux de vrais positifs (TVP) et de faux positifs (TFP) obtenus.

Tableau 11: Matrices de confusions comparées en termes de TVP et TFP. Les valeurs sont en pourcentage.

Vraie classe	Classe prédite SVM						Classe prédite KNN						Classe prédite RF					
	1	2	3	4	TVP	TFP	1	2	3	4	TVP	TFP	1	2	3	4	TVP	TFP
1	44	40	7	9	44	56	55	25	8	12	55	45	58	27	6	9	58	42
2	7	84	1	7	84	16	9	80	2	9	80	20	6	87	1	6	87	13
3	14	24	35	27	35	65	14	15	43	28	43	57	12	18	43	27	43	57
4	7	25	12	57	57	43	8	15	16	62	62	38	7	17	14	62	62	38

L'analyse comparative des trois matrices révèle qu'à l'exception des classes 3 (LH) et 4 (HH/HL) pour lesquelles les TVP sont égaux respectivement à 43 et 62 % pour le KNN et le RF ; le modèle RF présente les meilleurs pourcentages de TVP pour toutes les classes, et *a fortiori*, les TFP les plus bas. Conformément aux observations faites sur le nombre de points bien classés, la classe 2 obtient pour tous les modèles, le meilleur TVP. Ce pourcentage est de 87 % pour le RF, de 84 % pour le SVM et de 80 % pour le KNN. De même, en dehors des TVP de la classe 2, le modèle KNN obtient des scores plus élevés pour les classes 1, 3 et 4 que ceux du SVM.

Le tableau 12 présente les matrices de confusions comparées en termes de valeurs prédictives positives (VPP) et négatives (VPN).

Tableau 12 : Matrices de confusions comparées en termes de VPP et VPN. Les valeurs sont en pourcentage.

	Classe prédite SVM				Classe prédite KNN				Classe prédite RF			
Vraie classe	1	2	3	4	1	2	3	4	1	2	3	4
1	58	16	15	9	60	12	14	10	68	12	11	8
2	19	67	5	14	19	76	7	16	13	75	4	11
3	13	6	52	17	11	5	49	16	9	5	55	17
4	10	11	28	60	9	8	30	59	9	8	29	64
Valeur prédictive positive	58	67	52	60	60	76	49	59	68	75	55	64
Valeur prédictive négative	42	33	48	40	40	24	51	41	32	25	45	36

La valeur prédictive positive est la probabilité qu'un modèle classe correctement une cible minérale, lorsque la prédiction est vraie. *A contrario*, la valeur prédictive négative est la probabilité qu'un modèle ne classe pas correctement une cible minérale, lorsque la prédiction est erronée. L'analyse comparative des matrices en termes de VPP et VPN révèle qu'à l'exception de la classe 2 (LL), le modèle RF obtient les plus fortes valeurs de VPP et inversement les plus faibles valeurs de VPN pour les classes 1, 3 et 4.

Comparativement, pour les mêmes classes (1, 3 et 4), les VPP du SVM sont plus élevées que celles du KNN. Au niveau de la classe 2 (LL), le modèle KNN obtient un VPP légèrement supérieur au RF. Mais le VPP du RF reste supérieur au VPP du SVM. Il ressort de cette analyse comparative que de façon globale, le modèle RF est le meilleur prédicteur en termes de VPP pour les quatre types de cibles minérales. Cependant, pour la classe 2, les performances du KNN équivalent à celles du RF.

Le tableau 13 présente les valeurs des F-scores calculées pour chaque meilleur modèle.

Tableau 13 : Valeurs en pourcentage des F-score comparés.

Vraie classe	F-score SVM	F-score KNN	F-score RF
1	50	58	63
2	74	78	81
3	42	46	48
4	58	60	63

L'analyse des résultats de valeurs F-score obtenus pour chaque modèle, confirme que le modèle RF est réellement plus précis sur l'ensemble des quatre classes, que le SVM et le KNN. À son tour, le KNN est plus précis sur l'ensemble des classes que le SVM. De même, on constate que la classe 2 est la mieux classée par tous les modèles avec des F-scores de 81, 78 et 74 % respectivement pour le RF, KNN et le SVM.

En résumé, les analyses comparatives croisées réalisées en termes de précision globale de classification, de valeurs d'AUC et des matrices de confusions ont révélé que le modèle RF est plus précis à 69 % que les KNN et SVM sur le jeu de données d'apprentissage.

Le RF obtient pour toutes les classes, les plus grands nombres d'observations bien classées. Le RF présente les meilleurs pourcentages de TVP et de VPP calculés pour toutes les classes de cibles minérales. En outre, la structure du jeu de données d'apprentissage semble se refléter sur le nombre d'observations bien classées et sur les valeurs d'AUC estimées. Plus la composition numéraire de la classe augmente, plus la valeur d'AUC et le nombre d'observations bien classées augmentent.

Il ressort de cela également que plus le modèle est précis, moins il est sensible à la structure numéraire du jeu de données d'apprentissage. Le modèle SVM semble être beaucoup plus sensible à la structure du jeu de données d'apprentissage que les modèles prédictifs RF et KNN. Aussi, conformément aux observations faites sur les valeurs d'AUC, le nombre de points bien classés et les valeurs de TVP et de VPP de la classe 2 sont nettement supérieures à celles des autres classes pour tous les modèles. Nonobstant ces légères différences, les modèles RF et KNN semblent être plus précis, plus performants et meilleurs prédictifs sur le jeu de données d'apprentissage que les SVM.

Cependant, obtenir de bonnes performances sur un jeu d'entraînement, ne garantit pas que le modèle soit assez robuste et capable de performer sur de nouvelles données sur lesquelles il n'a pas été formé.

4.4.4 Analyse de robustesse et validation des meilleurs modèles prédictifs

Cette section présente et analyse les résultats des tests réalisés avec les données de validation afin d'évaluer la robustesse des meilleurs modèles sur un jeu de données externes. La structure (ou composition numéraire) du jeu de données de validation utilisé est présentée dans le tableau 14 et les résultats des tests analysés dans les paragraphes suivants.

Tableau 14 : Composition numéraire du jeu de données de validation.

Jeu de données	Effectif	Classe1 (%)	Effectif	Classe2 (%)	Effectif	Classe3 (%)	Effectif	Classe4 (%)	Total
Total	789	21	513	14	1542	42	894	24	3709
Validation	172	23	308	42	90	12	172	23	742

La composition du jeu de données de validation est sensiblement identique à celle du jeu de données d'apprentissage. À savoir, 308 entités pour la classe 2, soient 42 % de l'effectif total, 172 entités pour les classes 1 et 4, soit 23 % chacun et 90 entités pour la classe 3, soit 12 %. De ce fait, il est assez représentatif du jeu d'entraînement. Le tableau 15 présente les résultats des tests de robustesse et de validation effectués.

Tableau 15 : Matrices de confusions (en nombre et en pourcentage) des tests de validation

Matrice de confusion de validation du meilleur modèle RF (en nombre et en pourcentage)									
	Classe 1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)	Total	Classe1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)
Classe1 (AL)	109	40	9	14	172	63,37	23,26	5,23	8,14
Classe 2 (LL)	21	265	4	18	308	6,82	86,04	1,30	5,84
Classe 3 (LH)	9	18	33	30	90	10,00	20,00	36,67	33,33
Classe 5 (HH/HL)	9	29	26	108	172	5,23	16,86	15,12	62,79
Précision globale :					0,69				
Kappa :					0,56				
Matrice de confusion de validation du meilleur modèle KNN (en nombre et en pourcentage)									
	Classe1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)	Total	Classe1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)
Classe1 (AL)	104	36	16	16	172	60,47	20,93	9,30	9,30
Classe 2 (LL)	22	256	5	25	308	7,14	83,12	1,62	8,12
Classe 3 (LH)	11	15	35	29	90	12,22	16,67	38,89	32,22
Classe 4 (HH/HL)	8	22	33	109	172	4,65	12,79	19,19	63,37
Précision globale :					0,68				
Kappa :					0,54				
Matrice de confusion de validation du meilleur modèle SVM (en nombre et en pourcentage)									
	Classe1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)	Total	Classe1 (AL)	Classe 2 (LL)	Classe 3 (LH)	Classe 4 (HH/HL)
Classe1 (AL)	84	60	5	23	172	48,84	34,88	2,91	13,37
Classe 2 (LL)	21	265	4	18	308	6,82	86,04	1,30	5,84
Classe 3 (LH)	9	31	21	29	90	10,00	34,44	23,33	32,22
Classe 4 (HH/HL)	9	43	20	100	172	5,23	25,00	11,63	58,14
Précision globale :					0,63				
Kappa :					0,46				

Les résultats sont présentés en termes de précision globale, de valeurs d'AUC, de matrices de confusions (en nombre et en pourcentage) et de coefficients kappa.

Les résultats confirment que le modèle RF est légèrement plus précis et plus robuste (à 69,41 % pour un kappa de 0,56) que le modèle KNN. Ce dernier est précis à 67,92 % pour un kappa de 0,54. Le modèle SVM est précis à 63,34 % avec un kappa de 0,46.

Les valeurs des coefficients kappa obtenus ont été analysées en fonction de la littérature. Les seuils d'évaluation de bon accord des coefficients kappa peuvent changer d'un domaine ou d'une question à l'autre. Sur l'échelle proposée par Landis *et al.* (1977), une valeur proche de 1 indique un accord très fort, les valeurs comprises entre 0,6 et 0,8 un accord fort et les valeurs comprises entre 0,4 et 0,6 un accord modéré. Les modèles prédictifs RF et KNN obtiennent un kappa très proche de 0,6 et indique un accord fort. Le modèle SVM obtient un kappa compris entre 0,4 et 0,6 et indique un accord modéré.

Dans le domaine des sciences géologiques prédictives, plusieurs études ont donné des résultats similaires. Les travaux antérieurs de Kereszturi *et al.* (2018), portant sur la cartographie géologique de l'altération hydrothermale des volcans, à partir de données de classe de sol issues d'images multicateurs (*Sentinel-2* et *Landsat 8* OLI + TIR) ont utilisé une approche de modélisation par forêt aléatoire pour classer quinze (15) types de couvertures de sol.

Les résultats obtenus ont permis de fournir une carte géologique avec une précision globale de 69 % et un kappa de 66 %. D'autre part, les travaux de Miraki *et al.* (2019) pour la cartographie du potentiel des eaux souterraines de la plaine de Qorveh-Dehgolan, dans la province du Kurdistan (Iran) ont montré qu'un nouveau modèle hybride réalisé avec un RF présentait des performances plus élevées en termes de sensibilité, de spécificité et de précision, que les classificateurs basés sur une régression logistique et des Réseaux Bayésiens Naïfs. Ses résultats ont montré que le modèle RF fournit un kappa de 79 % pour les données d'apprentissage et 66 % pour les tests réalisés sur les données de validation.

Des valeurs de précision globale de classification de 68 et 69 % et des coefficients kappa de 0,5 et 0,6 ont été obtenues respectivement avec les modèles prédictifs KNN et SVM sur un nouveau jeu de données et ont été jugées satisfaisantes.

Sur la base de cette littérature, nos résultats peuvent être jugés satisfaisants. Les modèles RF et KNN sont par conséquent plus précis et plus performants que le modèle SVM. Cependant, force est de constater que la validation des performances des modèles RF et KNN montre des résultats très comparables en termes de valeurs kappa obtenues, ce qui témoigne (une fois de plus) de la robustesse de ces deux algorithmes utilisés. D'un autre côté, les faibles performances des algorithmes SVM, peuvent être dues à un faible nombre de données utilisées pour leur entraînement, car le nombre de paramètres utilisés linéairement pour la formation des algorithmes SVM est lié au nombre de données d'apprentissage (Cortes *et al.*, 1995). Ces classificateurs nécessitent généralement une très grande base de données d'apprentissage et sont des modèles discriminants qui tentent de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les données des classes. La maximisation de la marge étant une méthode de régularisation qui réduit la complexité de ces classificateurs.

4.5 Analyse des nouvelles cartes prédictives minérales

En Côte d'Ivoire il existe plusieurs ébauches de cartes prédictives de potentialités minérales (en métaux lourds) établies par des géologues et utilisées à des fins de prospections plus approfondies. Ces cartes sont pour la plupart réalisées en appliquant différentes méthodes de géotraitement (de krigeage ou de cokrigeage) à un ensemble de données géochimiques couplées à différents critères de preuves de minéralisation (géomorphologiques, structuraux, lithologiques, géophysiques, etc.) ou couplées à des données issues du traitement d'images optiques ou radar.

Il s'agit notamment d'ébauches de cartes prospectives réalisées à l'échelle de gisements connus peu ou mal prospectés. Ces cartes sont pour la plupart utilisées pour le ciblage de nouvelles zones à prospecter de façon plus approfondie. Au nombre de celles-ci, figurent les cartes de potentialités minérales des placers diamantifères de Tortiya (région de Séguéla au nord de la Côte d'Ivoire) et des ébauches de cartes géochimiques réalisées dans le cadre de ce projet d'exploration (Projet coltan d'Issia), d'où sont extraites les données de puits utilisées.

Pour des raisons de confidentialités inhérentes aux compagnies minières privées, ces ébauches de cartes prospectives n'ont pu être présentées dans ce mémoire. Cependant, l'approche proposée par notre étude est (à ce jour) l'une des premières en la matière en Côte d'Ivoire, à utiliser la puissance prédictive des algorithmes de classification automatique, pour prédire sur de vastes étendues géographiques peu ou mal explorées (et à des taux satisfaisants) de nouvelles zones favorables à la formation de gîtes secondaires de minéraux lourds (coltan) de type placers.

Les figures 42 et 43 présentent respectivement les cartes de potentialités minérales en coltan réalisées pour les sites de Brokoua – Makua et pour le site de Bémadi.

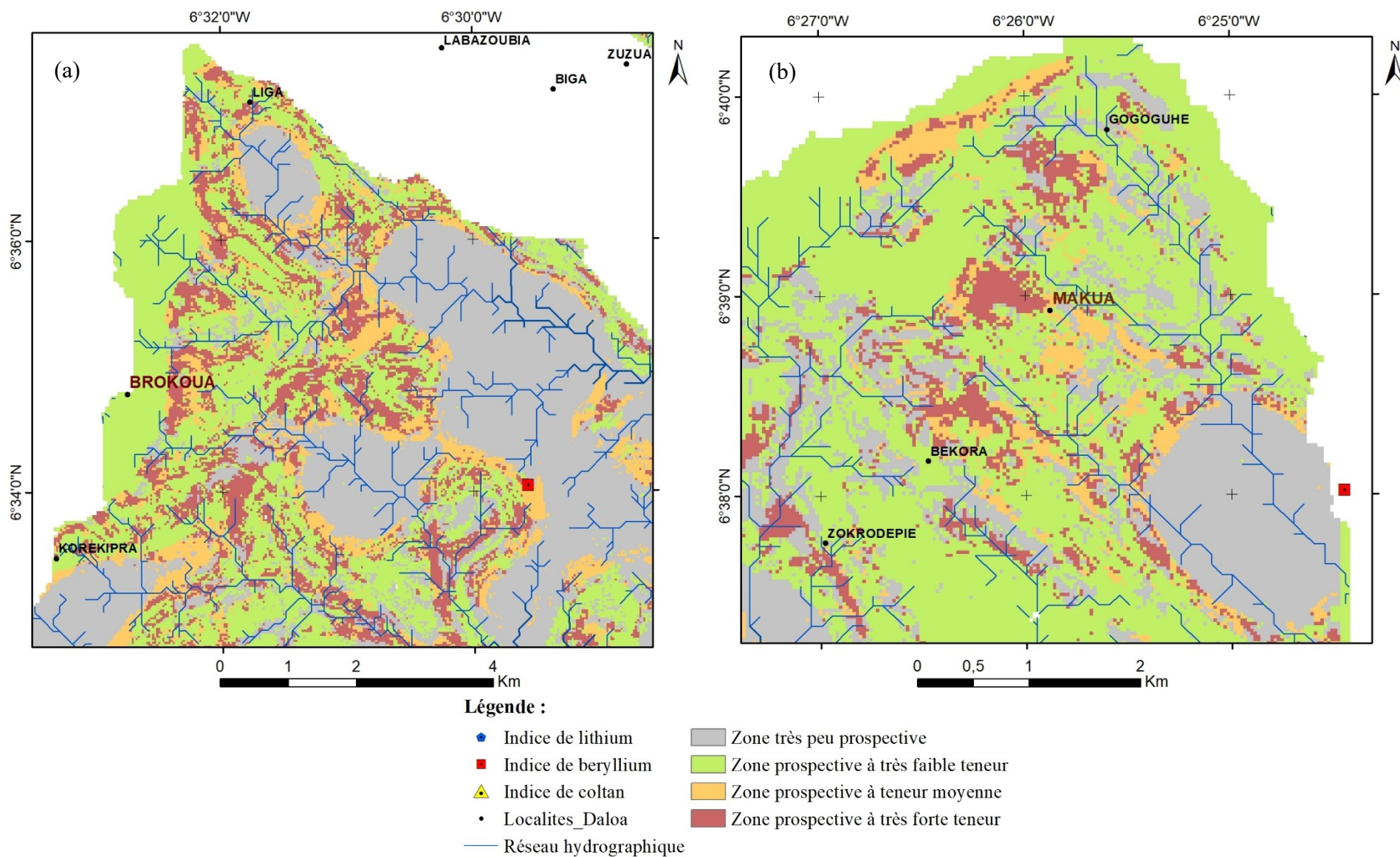


Figure 42: Cartes des potentialités minérales de coltan : (a) Secteur de Brokoua - (b) Secteur de Makua.

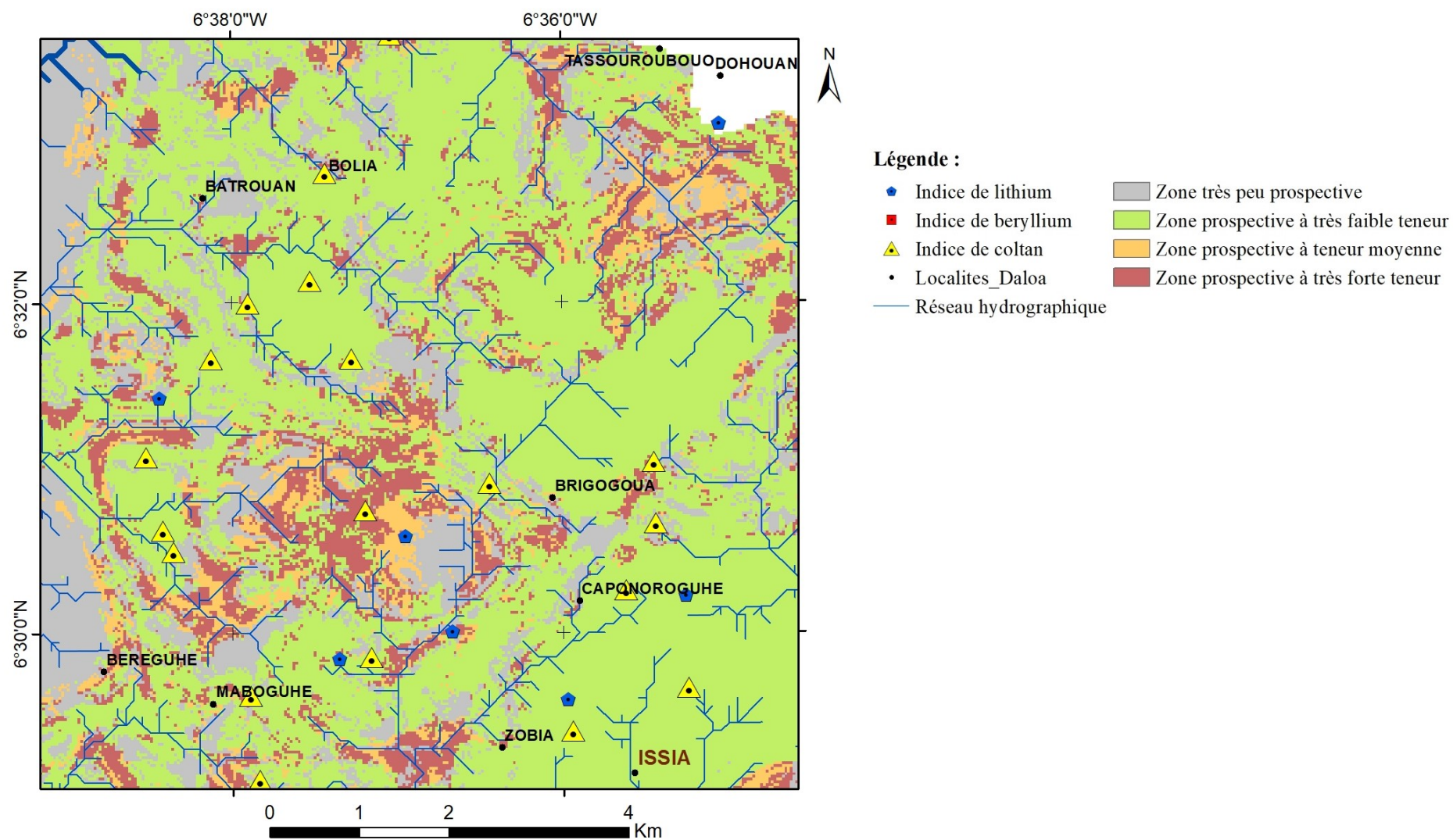


Figure 43: Cartes des potentialités minérales de coltan : Secteur de Bémadi.

Les zones en gris clair correspondent aux zones potentiellement peu prospectives. Ces zones représentent peu d'intérêts, au sens de la prospectivité de nouvelles cibles minérales économiquement rentables. Les minéralisations des graviers dans ces zones sont susceptibles de varier considérablement à l'échelle locale. Ce sont des zones qui ne présentent aucune agrégation spatiale statistiquement significative à l'échelle des sites étudiés. En revanche, l'analyse de la distribution spatiale des nouvelles zones prospectives, met en évidence deux catégories de zones potentiellement favorables pour une concentration de graviers minéralisés à valeurs économiques.

La première catégorie est composée de zones favorables à la concentration de graviers de teneurs moyennes (en orange) et de teneurs fortes (en rouge). Ces surfaces correspondent en majorité aux zones d'interfluves et de bordures de rivières (ou de marigots) mais également, aux zones situées à proximité d'affleurements rocheux de pegmatites granitiques minéralisés. Sur le plan hydromorphologique et structural, ces surfaces occupent des niveaux topographiques surélevés et correspondent aux zones soumises à un faible drainage et marquées par une densité de fracturation moyenne à élevée.

La deuxième catégorie se compose de zones potentiellement favorables à la concentration de graviers minéralisés à faibles teneurs (en vert). Il s'agit des zones occupant des niveaux topographiques plus faibles, soumises à d'importants processus de drainage et caractérisées par une densité de fracturation plus faible. Elles sont très étendues et occupent la plus grande majorité des flats des sous-bassins versants étudiés.

Aux plans géologique et minéralogique, la littérature et les travaux de terrain ont prouvé que la plupart des graviers minéralisés en coltan et des indices minéraux sont localisés au niveau de la zone de contact (schistes – granites) définie par Cruys (1965); notamment, entre les formations de granite à biotite subordonnée, de schistes sériciteux et chloriteux et de schistes argileux (Allou, 2005).

À l'échelle du Complexe géologique, la localisation de plusieurs indices minéraux regroupés sur un espace géographique, est un bon indicateur de présence de dépôts potentiels.

Le tableau 16 montre la distribution en termes de superficies et de pourcentages des nouvelles zones cartographiées, selon la zone de contact de Cruys.

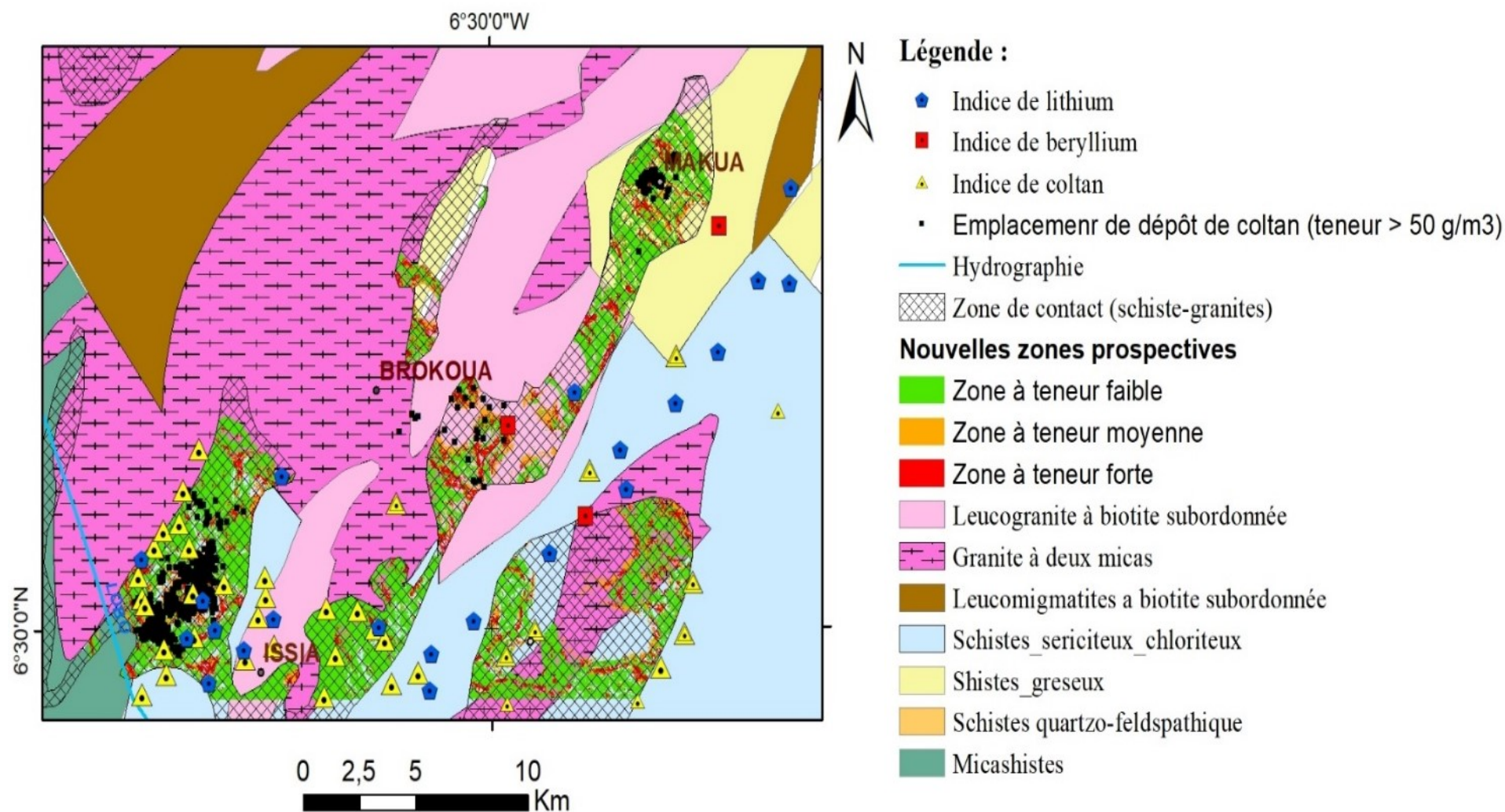
Tableau 16: Répartition en pourcentage des zones prospectives dans la zone de contact.

Zone de contact schistes - granites	Superficie totale (km ²)	Nature de la zone							
		Peu prospective	(%)	Prospective à teneur faible	(%)	Prospective à teneur moyenne	(%)	Prospective à teneur forte	(%)
Couverture spatiale (km ²)	183,59	58,13	32	79,02	43	10,62	6	17,67	10

L'analyse de ce tableau montre que la zone de contact identifiée par Cruys (1965) s'étend sur près de 184 km². Les zones peu prospectives en occupent 32 %, tandis que les zones prospectives à teneur faible en occupent 43 %. Les zones prospectives à teneurs moyennes et fortes occupent respectivement 6 et 10 %.

Ce résultat est en adéquation avec les observations faites sur le terrain, lors, des différents travaux de prospection menés, dans la région nord du Complexe géologique Hana-Lobo.

La carte de la figure 44 présente l'occupation spatiale des différentes zones prospectives identifiées, en fonction de la zone de contact schistes – granites de Cruys. Elle situe notamment l'emplacement des dépôts exploitables (teneurs supérieures à 50 g/m³) connus de la région nord du Complexe géologique.



L'analyse et l'interprétation de cette carte, confirme bien que 35 indices minéraux de coltan connus (en jaune) sont localisés au niveau d'une zone de contact entre des leuco granites à biotites subordonnées et des schistes sériciteux – chloriteux. Sur ces 35 indices connus, 21 sont situés de part et d'autre et à proximité des bordures de la zone de contact cartographiées par Cruys en 1965, soit 60 % des indices de coltan connus.





Nos résultats mettent ainsi en évidence la présence de plusieurs zones potentiellement prospectives pour des dépôts économiques, au niveau des emplacements de ces 21 indices de coltan. De plus, l'observation de la figure 29 montre que la grande majorité des dépôts exploitables de coltan (teneur supérieure à 50 g/m³), en noir sur la carte, coïncident avec la zone de contact de Cruys. Ce qui suggère une certaine influence de ces zones de contacts schistes – granites sur la distribution spatiale des dépôts de coltan dans les secteurs étudiés.

4.6 Le géoportail interactif de l'information géominière

L'outil d'aide à la recherche d'information géominière spatiale a été baptisé « Infominci ». Il s'agit d'un géoportail qui permet de visualiser, de façon démocratique (sans logiciel SIG nécessaire) et au plus près de la source, l'information minéralogique, géologique, géophysique, et administrative. Il est utilisable pour collecter de l'information complémentaire et guider des travaux de prospection approfondis réalisables sur les nouvelles zones potentiellement prospectives ciblées. De façon pratique, l'application met à la disposition de l'utilisateur (exemple : le prospecteur géologue), une couche interrogeable dénommée « Indice documenté », qui fait référence à un sommaire des principales thématiques abordées dans les rapports d'exploration de plusieurs indices minéraux sur toute l'étendue de la Côte d'Ivoire. L'ensemble de l'information géospatiale est accessible depuis un téléphone intelligent, une tablette ou un ordinateur de bureau, via une page *Web* disponible et consultable au lien hypertexte suivant :

http://igeomedia.com/~kassi/gabarit_Web_infominci/.

La figure 45 présente une vue générale et plusieurs vues des différentes sections du géoportail. Le paragraphe suivant en décrit les composantes ainsi que leur fonctionnement. La vue générale présente en son centre, un espace d'affichage cartographique convivial et interactif qui permet de visualiser les couches thématiques sélectionnées. Il s'agit notamment d'une image satellite et d'une couche *open street map* de haute résolution spatiale, utilisables comme carte de fond. La carte de prospectivité du coltan de la zone d'étude (en (A)) a également été intégrée au SIG *Web*.

Les icônes  en haut à gauche (B) permettent de réaliser un zoom avant et arrière. L'icône de menu rétractable  en (C), permet d'afficher et de masquer les onglets (Couches, Description du rapport, Présentation et contacts et Légende) afin d'offrir une vue cartographique beaucoup plus dégagée. Les icônes de navigation de la page *Web* , en haut à droite ((D) et (E)), permettent de réaliser les fonctions classiques de vue initiale, de déplacement, de sélection de zones et d'interrogation de couche. En outre, l'icône  en (E) permet un affichage cartographique en plein écran. Il est à noter que cette même partie est le lieu d'affichage des coordonnées géographiques en degrés décimaux de la position du curseur.

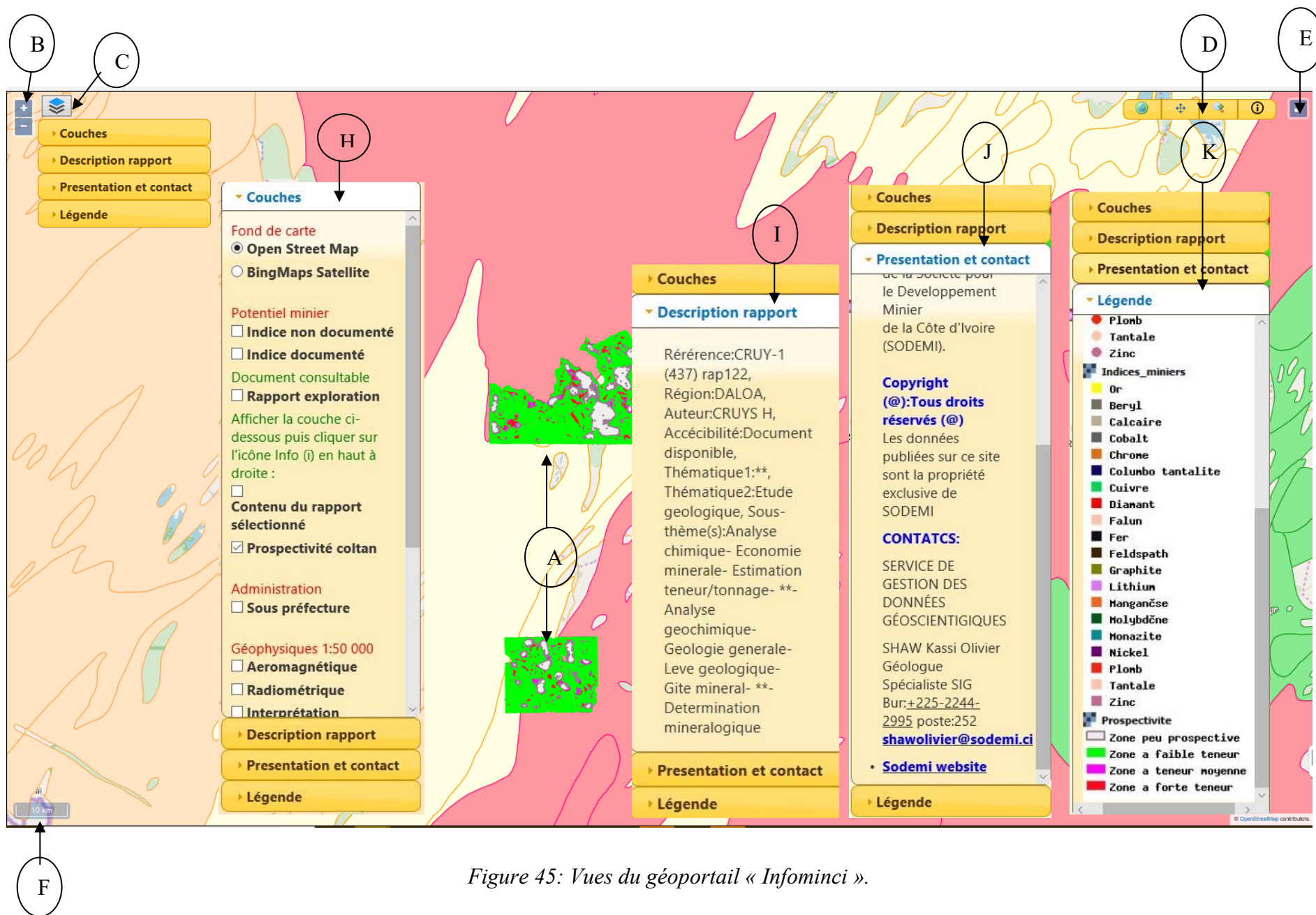



Figure 45: Vues du géoportail « Infominci ».

Dans la partie située en bas à gauche (F), figure l'échelle d'affichage cartographique. S'agissant des vues d'onglets de menus de haut en bas, l'onglet « Couches » en (H) permet de sélectionner, d'afficher et de masquer les différentes couches d'information géospatiale et de sélectionner à l'aide de l'icône  un indice et son rapport, pour en obtenir une description sommaire. En effet, pour chaque indice documenté sélectionné, l'onglet « Description du rapport » (I) apparaît et présente le résultat, en termes d'auteur, d'année de publication, de matières et de thématiques abordées dans chaque œuvre.

L'onglet suivant (J) présente de façon brève l'application ainsi que les contacts de l'administrateur. Enfin, le dernier onglet en bas (K), donne accès aux différentes légendes cartographiques nécessaires pour interpréter les différentes couches géominières (indices miniers et prospectivité du coltan). Les données de couches ont été stockées dans le SGBRS suivant les normes de l'OGC. À ce titre, elles sont consultables par requêtes WMS (s'agissant des couches au format raster) et par requêtes WFS (s'agissant de la couche interrogeable (indices documentés)). Cela garantit l'interopérabilité du SIG *Web*.

Le géoportail minier ainsi développé, peut être utilisé pour diffuser de façon démocratique et au plus près de la source, l'information géominière nécessaire pour optimiser une planification d'exploration des nouvelles cibles minérales identifiées dans l'étude. Il peut être utilisé également, pour guider des travaux de recherche minière plus approfondie si nécessaire. Enfin, ce prototype est susceptible d'être implémenté dans une application mobile, de sorte à vulgariser son utilisation au bénéfice du géologue prospecteur sans connaissances avancées du champ SIG.

Cette partie du travail vise à valider et discuter les résultats de l'étude au regard de la littérature et des travaux antérieurs réalisés sur notre zone d'étude.

4.7 Cibles minérales

Dans cette étude, nous avons utilisé la technique d'analyse statistique d'agrégation spatiale de Moran pour expliquer le mode de distribution de 3709 échantillons de graviers minéralisés issus des travaux de terrain. Les résultats de cette analyse statistique ont montré qu'à l'exception de 789 points distribués (à l'échelle de 30 m) de façon aléatoire, le mode de distribution géostatistique des données de prospection minérale est statistiquement significatif à 96 % pour discriminer trois types de cibles minérales prospectives.

Les échantillons utilisés sont significativement regroupés en grappes spatiales et peuvent être discriminés et classifiés selon une valeur seuil de 44 g/m³ pour représenter les dépôts connus identifiés par plusieurs anciens travaux (Rombach, 1960 ; Cruys, 1965 ; Adam, 1968 ; Papon, 1971 ; Feybesse *et al.*, 1990 ; Vidal *et al.*, 1992 ; Allou, 2005). Par conséquent, le mode d'agrégation spatiale de données de puits minéralisés et les valeurs des teneurs moyennes (en concentrés de minéraux lourds) peuvent être utilisées pour discriminer, représenter et cartographier différents types de cibles minérales prospectives dans le cadre d'une étude de CPM.

Cependant, 789 échantillons rejettent l'hypothèse nulle et ne présentent aucune agrégation spatiale statistiquement significative à 96 %. Une telle dissemblance observée entre ces échantillons et au sein d'un espace continue, peut figurer l'existence de plusieurs zones d'interruptions (à l'échelle de 30 m) des phénomènes sous-jacents contrôlant les processus d'érosion, de transport et de dépôt des minéraux lourds plus ou moins minéralisés en coltan (0 à 1126 g/m³). Des travaux complémentaires à l'échelle de ces zones de distribution géostatistique sont à envisager en vue de mettre en évidence (de façon plus détaillée) la présence d'éventuelles sous-zones anormales, favorables à la concentration de graviers minéralisés en coltan avec des teneurs exploitables (supérieures à 44 g/m³). Des travaux de fonçage de puits complémentaires peuvent être entrepris à cet effet au niveau de ces zones potentiellement peu prospectives à l'échelle régionale. Il s'agira d'étudier à une échelle

plus réduite (< 30 m) la variation spatiale des facteurs hydromorphologiques et structuraux considérés. Puis, de prélever de nouveaux échantillons suivant une maille inférieure à 25 m et réaliser des études minéralogique et géochimique complémentaires plus approfondies, pour tenter d'expliquer ces conditions particulières de dépôts ou d'occurrences minérales dispersées.

4.8 Prédicteurs spatiaux

Ce mode de distribution géostatistique observé implique l'existence de facteurs sous-jacents à l'œuvre et susceptibles d'influencer et de contrôler la distribution géostatistique des cibles minérales sur la zone d'étude. Nous avons pu extraire à partir des données géospatiales, plusieurs paramètres hydromorphologiques et structuraux. Cela nous a permis d'analyser leur capacité à influencer ou contrôler la localisation des dépôts secondaires au sein de la zone d'étude. L'analyse de nos résultats a relevé plusieurs observations pertinentes.

Au plan morpho structural, nous avons mis en évidence un système de fracturation de directions variées (NS (NN0-10), NE-SW (NN20-50), EW (NN90-110) et SE-NW (NN120-150)). Ce système est en accord avec les directions régionales identifiées dans des travaux de cartographie structurale réalisés au centre-ouest de la Côte d'Ivoire (Koudou *et al.*, 2010, 2014) et plus particulièrement au niveau du bassin de la Lobo (Yao *et al.*, 2016). Ces directions mises en évidence s'intègrent bien au système de déformation tectonique régional ayant donné naissance aux plis droits et isoclinaux de la zone de contact (schistes-granites) mentionnée par Triboulet *et al.* (1998) et Vidal *et al.* (2009).

Nous avons également observé que les emplacements d'affleurements connus de pegmatites granitiques minéralisés (au niveau du site de Bémadi) coïncident avec une zone à forte densité de fracturation, et que la majeure partie des cibles minérales discriminées sont logées soit au cœur (site de Bémadi) ou à la périphérie (site de Makua et Brokoua) de ces zones fortement fracturées. L'influence des zones de densité de fracturation moyenne à élevée, sur la localisation des cibles minérales à très fortes teneurs au sein des micaschistes (zone de contact métamorphique) et à proximité d'une zone de concentration d'affleurements de pegmatites granitiques minéralisés en Ta/Nb, est en accord avec le

modèle génétique proposé par Allou (2005). Ce modèle stipule que la mise en place des pegmatites crée de fortes anisotropies de compétences dans les micaschistes dans lesquels elles se forment. Les pegmatites de plus en plus stériles se forment à la périphérie du granite de type Issia.

Sur le plan hydrologique, l'observation des paramètres de surfaces nous a permis de mettre en évidence différentes zones de drainages. Les zones d'interfluves et les zones d'irrégularités topographiques (crêtes, flancs ou sommets) sont soumises à une amplitude de flux élevée et les zones de flats et de vallées sont soumises à une amplitude de flux moins élevée. Nous avons utilisé ces facteurs sous-jacents, dans la modélisation prédictive, comme prédicteurs spatiaux pour traduire l'influence de ces paramètres sur la formation de dépôts secondaires de minéraux lourds. Deux catégories de nouvelles zones prospectives ont pu ainsi être prédites et cartographiées.

4.9 Nouvelles zones prospectives

Nouvelles zones prospectives à teneurs moyenne et forte :

La première catégorie est constituée de zones favorables à la concentration de graviers de teneurs moyennes à fortes. Ces nouvelles zones correspondent en majorité aux zones d'interfluves et de bordures de rivières ou de marigots, mais également aux zones situées à proximité d'affleurements rocheux de pegmatites granitiques.

L'influence de ces facteurs hydromorphologiques sur la distribution spatiale des cibles très minéralisées suggère que cette zone est peu favorable aux processus d'altération météorique des roches hôtes primaires et au drainage de minéraux lourds. Cela signifie que ces minéraux lourds peu érodés, qui n'ont pas connu un long déplacement au cours des ruissellements successifs, s'accumulent préférentiellement, dans les colluvions et les éluvions à proximité des minéralisations primaires.

Ce constat confirme les résultats antérieurs obtenus lors de l'étude minéralogique des placers de la région d'Issia (Allou, 2005), dont l'étude granulométrique des graviers a démontré que le colombo-tantalite libéré sur les crêtes et les flancs des interfluves est en général de grande taille et de forme anguleuse. Une telle morphologie des grains de coltan est le résultat d'une faible corrosion mécanique de ces minéraux induite par un faible

processus de transport. De plus, ses travaux de terrain ont montré qu'au niveau des interfluves, se développaient des sols latéritiques et des graviers à fortes teneurs en minéraux colombo-tantalifères, vu la grande quantité de pegmatites minéralisées et la conservation des dimensions des grains de coltan échantillonnés. Nos résultats viennent également confirmer les travaux de Rombach (1960) et Cruys (1965) qui ont mis en évidence, des gisements de concentrations détritiques en minéraux de Tantale (Ta) et de Niobium (Nb), issues de la désagrégation de roches primaires et situées aux sommets et sur les flancs des collines situées en amont des zones alluvionnaires des placers minéralisés.

D'autre part, la forme de ces nouvelles zones prospectives à teneurs moyennes et fortes est généralement étroite et allongée au niveau des bordures de rivières, de marigots et un peu plus étalée au niveau des zones d'interfluves, formées durant les grandes phases d'alluvionnement qu'a connu la région. Cela confirme que ces zones d'anfractuosités exercent une influence sur la distribution spatiale de ce type de dépôt. Ces zones constituent bien, des zones de pièges naturels de minéraux lourds fixes dans l'espace et dans le temps. Nos observations sont confirmées par les travaux de Jacob et *al.* (1999). Leurs travaux portaient sur la formation des gisements diamantifères de type placers, de la rivière Orange en Afrique du Sud. En effet, ils ont mis en évidence trois types de pièges naturels constitués d'irrégularités topographiques fixes dans l'espace et favorables à l'accumulation au même endroit, de minéraux lourds de diamants à valeurs économiques.

De ce fait, nos résultats ont démontré que le faible potentiel à transporter, fragmenter et éroder les roches hôtes pegmatitiques fortement minéralisées, peut être utilisé comme un indicateur spatial pour la recherche de nouvelles zones favorables à la concentration de ces pegmatites minéralisées peu altérés, mais aussi, comme critère spatial de recherche de nouveaux dépôts de type placers moyennement et fortement minéralisés.

Nouvelles zones prospectives à teneur faible :

La deuxième catégorie, composée de zones potentiellement favorables à la concentration de graviers minéralisés à faible teneur, coïncide avec des zones occupant des niveaux topographiques plus faibles. Ces zones sont soumises à d'importants processus de drainage.

L'influence des paramètres hydromorphologiques sur la distribution spatiale de ce type de dépôt confirme les observations de terrain obtenues lors de l'analyse des facteurs et paramètres explicatifs de la dynamique de distribution et de genèse des dépôts de coltan de la région d'Issia (Allou, 2005). Concernant le secteur de Bémadi, ses travaux ont montré que les teneurs relativement plus faibles, observées dans les alluvions de *flats* et de vallées, sont la cause conjuguée d'une plus faible présence de pegmatites et d'une action destructrice des forts courants fluviaux. Ces minéraux transportés sur de longues distances, se fragmentent en particules de colombo tantalite plus fine, se dispersent, s'étalent et se reconcentrent dans les alluvions. Ce mode de transport et de dépôt de minéraux lourds trouve une explication suivant la loi de Stokes (Bouladon, 1964). Celle-ci stipule qu'il y a concurrence entre la gravité et la vitesse de l'eau. Ainsi, un minéral lourd se déposera d'autant plus loin que le courant est plus fort (Allou, 2005).

Enfin, sur le plan géologique, nos résultats ont montré que les zones de contacts schistes-granites sont un bon indicateur de prospection de dépôts à teneur moyenne et forte. Ces zones renferment 60 % des indices liés aux minéralisations de pegmatites granitiques de type Li -Be - Ta/Nb. Par conséquent, des critères de preuves hydromorphologiques et structuraux peuvent être utilisés comme prédicteurs spatiaux, dans une étude de CPM basée sur l'utilisation d'algorithmes de classification par apprentissage automatique, pour modéliser de façon prédictive la distribution géostatistique de nouvelles cibles minérales dans des régions d'intérêts.

4.10 Modèle prédicteur final

Dans cette étude, l'objectif a été de développer un modèle qui soit suffisamment complexe pour bien capturer la nature des données (en évitant un sous-apprentissage), et suffisamment simple pour éviter un surapprentissage. En somme, il fallait trouver un modèle intermédiaire avec à la fois un biais de prédiction faible et une bonne capacité à généraliser. La prise en compte de la technique de validation croisée (*5-fold*) dans les processus d'apprentissage des algorithmes a permis d'éviter d'éventuels biais liés au phénomène de surajustement lors de l'apprentissage.

Au niveau de l'apprentissage automatique, les résultats obtenus ont démontré que le meilleur modèle *Random Forest* est précis à 70 %, pour classer correctement une grande proportion de cibles minérales pertinentes et proposées dans les jeux de données. Le meilleur modèle RF est aussi précis que le meilleur modèle KNN. Tous ces deux modèles sont plus précis que le SVM.

Au niveau de la sensibilité et de la spécificité des modèles, les résultats ont démontré que tous les trois meilleurs modèles prédicteurs (KNN, SVM et RF) avaient une meilleure capacité à classer et prédire les entités de la classe 2 (LL), quand leurs performances diminuaient comparativement avec les classes 3 (LH), 4 (HH/HL) et 1 (AL). Cette légère baisse de compétence pourrait s'expliquer par une distribution hétérogène des modalités de classes dans le jeu de données d'apprentissage. Ce qui pourrait exercer une influence sur l'apprentissage de ces algorithmes. Cependant, les partitions des jeux de données de validation et d'apprentissage ont été créées de façon aléatoire, pour satisfaire à une rigueur scientifique et garantir une certaine robustesse des modèles de sortie, face à différents scénarii d'apprentissage. Dans le monde réel et particulièrement dans le domaine de la prédictivité minérale, une dissymétrie dans les jeux de données, n'est pas l'exception, mais la règle. Dans le sens où, les minéralisations sur de vastes étendues géographiques, peuvent être considérées comme des événements rares (Xiong et *al*, 2018). Aussi, il faut noter que l'apprentissage, à partir de données dont la distribution des modalités de classe est très éloignée d'une distribution uniforme, est une situation relativement fréquente dans les cas des classifications par apprentissage automatique, en particulier dans les études de CPM.

Le modèle prédicteur final RF utilisé, satisfait au concept de compromis biais-variance. Car, la variation des valeurs du paramètre de profondeur de l'arbre, appliquée pour sélectionner le meilleur modèle prédicteur RF dans cette étude, a permis de contrôler le biais et la variance de cet algorithme d'apprentissage. De plus, le processus de formation du modèle final, avec la méthode d'ensachage (*bagging*) a permis de combiner de meilleurs classifieurs, ce qui a également contribué à réduire la variance. Par conséquent, notre algorithme final est plus complexe, plus avancé et plus robuste. Il combine les résultats de nombreux apprenants faibles ou simples en un modèle d'ensemble de haute qualité. La comparaison de nos résultats obtenus en fonction de ceux mentionnés dans plusieurs études

antérieures de CPM (Carranza *et al.*, 2015, 2016; Hariharan *et al.*, 2017; Rodriguez-Galiano *et al.*, 2015) atteste que le modèle RF est généralement plus performant et plus robuste que les KNN et les SVM. Tenant compte de cette discussion, le modèle prédictif final RF est un classifieur assez puissant et assez robuste pour prédire, à des taux satisfaisants, de nouvelles zones prospectives dans le cadre de cette étude de CPM.

Cette partie du travail vise à valider les hypothèses de recherche.

Nos résultats démontrent que l'utilisation de la technique d'analyse d'agrégation spatiale de Moran est une technique valide, pour discriminer différents types de cibles minérales colombo tantalifères, dans un contexte de données d'exploration minière incomplètes. Nous avons également cartographié plusieurs critères de preuves hydromorphologiques et structuraux, susceptibles d'expliquer, d'influencer ou de contrôler les processus de formation des dépôts détritiques de type placers étudiés. Nos résultats confirment également à 70 % que ces facteurs peuvent être utilisés comme prédictifs spatiaux pour modéliser la distribution géostatistique de dépôts connus et prédire à des taux satisfaisants la prospectivité minérale de nouvelles cibles d'intérêts, tout en évitant une interprétation potentiellement biaisée faite par un géologue.

Notre étude est l'une des premières en Côte d'Ivoire à utiliser des techniques de CPM par apprentissage automatique pour cartographier et prédire, à partir d'observations et de données de prospection géologique incomplètes, de nouvelles zones potentiellement favorables aux dépôts de coltan de type placer. Nos résultats contribuent donc, à la mise à jour des connaissances géoscientifiques acquises sur la formation et la distribution spatiale des gîtes secondaires des minéraux lourds (coltan) dans un contexte de climat tropical humide.

CHAPITRE V. CONCLUSIONS ET RECOMMANDATIONS

Dans un contexte de données d'exploration minière incomplètes et de dimensionnalité spatiale élevée, notre étude a montré d'une part que les algorithmes de classification automatique (KNN, SVM et RF) sont performants pour modéliser de façon prédictive les relations spatiales complexes et non linéaires existantes entre des données de prospection géologique et différents critères spatiaux de preuves hydromorphologiques sous-jacents. Les modèles prédictifs RF (30 divisions et 100 apprenants) et KNN (pondéré-euclidienne IDW) sont précis respectivement à 70 % et 66 %. Le modèle SVM (noyau gaussien fin) est précis à 63 %. Les modèles RF et KNN sont robustes à 70% pour un kappa de 0,6 et le modèle SVM à 64 % pour un kappa de 0,5. Ces résultats permettent d'identifier, à des taux satisfaisants, de nouvelles zones potentiellement favorables à la formation et à la prospection économique de nouveaux dépôts détritiques de coltan. D'autre part, notre étude a mis en évidence l'influence des processus hydromorphologiques et structuraux (érosion, drainage, système de fracturation et topographie) sur la distribution spatiale des différents types de cibles minérales identifiées, au sein des sous-bassins versants du Complexe géologique Hana-Lobo. Cela démontre que ces paramètres peuvent être utilisés comme critères de preuves pour modéliser et prédire de nouveaux emplacements favorables à la mise en place de dépôts détritiques, dans un contexte de minéralisation de type placer. Cependant, les résultats obtenus et les conclusions tirées ne valent que pour la partie nord du Complexe. Ils n'ont pu être validés sur les sites de campagne de terrain, du fait de l'insuffisance du nombre de puits forés. Ainsi, nous recommandons de réaliser des puits complémentaires au sud du Complexe géologique, en vue de valider les nouvelles zones potentiellement prospectives identifiées dans l'étude. Vu le potentiel de l'approche proposée, l'étude pourrait être étendue à tous les bassins versants de la Côte d'Ivoire susceptibles d'héberger une minéralisation pegmatitiques de type Niobium – Tantale – Béryllium – Lithium (Nb-Ta, Be, Li). Au niveau de la capacité prédictive de l'algorithme de classification RF, nous recommandons d'évaluer l'impact des nouvelles méthodes *Balanced Random Forest* et *Weighted Random Forest* sur la classification automatique d'un jeu de données de prospection minière déséquilibré. Dans le cadre de la préparation des travaux complémentaires de prospection sur les nouvelles cibles potentielles

cartographiées, nous recommandons l'utilisation du géoportail Infomici, développé dans l'étude, comme outil d'aide de collecte et d'analyse d'informations spatiales géominières disponibles ou à acquérir. Ce qui permettra de réduire considérablement les risques et les coûts liés à l'exploration. Enfin, cette étude sera l'objet d'un article scientifique destiné à être publié dans une revue scientifique cotée de géosciences.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abedi, M., Norouzi, G.-H. et Bahroudi, A. (2012) Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences*, vol. 46, p. 272-283.
- Abedi, M. et Norouzi, G.-H. (2012a) Integration of various geophysical data with geological and geochemical data to determine additional drilling for copper exploration. *Journal of Applied Geophysics*, vol. 83, p. 35-45.
- Abedi, M., Mostafavi Kashani, S. B., Norouzi, G. H. et Yousefi, M. (2017) A deposit scale mineral prospectivity analysis: A comparison of various knowledge-driven approaches for porphyry copper targeting in Seridune, Iran. *Journal of African Earth Sciences, GIS-based mineral potential targeting*, vol. 128, p. 127-146.
- Adam, H. (1968) Les pegmatites de la région d'Issia. Abidjan, 69 p.
- Adiri, Z., El Harti, A., Jellouli, A., Maacha, L., & Bachaoui, E. M. (2015). Lithological mapping using multispectral ASTER and Landsat 8 data in the Bas Drâa inlier, Moroccan Anti Atlas. In *Earth Resources and Environmental Remote Sensing/GIS Applications VI* (Vol. 9644, p. 96440L). SPIE. <https://doi.org/10.1117/12.2193775>
- Adiri, Z., El Harti, A., Jellouli, A., Lhissou, R., Maacha, L., Azmi, M., Zouhair, M. et Bachaoui, E. M. (2017) Comparison of Landsat-8, ASTER and Sentinel 1 satellite remote sensing data in automatic lineaments extraction : A case study of Sidi Flah-Bouskour inlier, Moroccan Anti Atlas. *Advances in Space Research*, vol. 60, n°11, p. 2355-2367.
- Ali, S. S., Nizamuddin, S., Abdulraheem, A., Hassan, M. R. et Hossain, M. E. (2013) Hydraulic unit prediction using support vector machine. *Journal of Petroleum Science and Engineering*, vol. 110, p. 243-252.
- Allou, B. A. (2005) Facteurs, paramètres, dynamique de distribution et genèse des dépôts de columbo-tantalite d'Issia, centre-ouest de la Côte d'Ivoire (phd). Thèse de doctorat, Université du Québec à Chicoutimi, Montréal, 352 p.
- Amer, R., Kusky, T. et El Mezayen, A. (2012) Remote sensing detection of gold related alteration zones in Um Rus area, Central Eastern Desert of Egypt. *Advances in Space Research*, vol. 49, n°1, p. 121-134.
- Arnould, M. (1961) Étude géologique des migmatites et des granites précambriens du

- Nord-Est de la Côte d'Ivoire et de la Haute-Volta méridionale. Mém. BRGM. Thèse de doctorat, Clermont-Ferrand, 174 p.
- Asadi, H. H., Porwal, A., Fatehi, M., Kianpouryan, S. et Lu, Y. J. (2015) Exploration feature selection applied to hybrid data integration modeling: Targeting copper-gold potential in central Iran. *Ore Geology Reviews*, vol. 71, p. 819-838.
- Anthoine, R. (1941) Les méthodes pratiques d'évaluation des gîtes secondaires aurifères appliquées dans la région de Kilo-Moto. (Congo belge) – Mémoire, Institut Royal Colonial Belge, 249 p.
- Balabin, R. M., Safieva, R. Z. et Lomakina, E. I. (2011) Near-infrared (NIR) spectroscopy for motor oil classification : From discriminant analysis to support vector machines. *Microchemical Journal*, vol. 98, n°1, p. 121-128.
- Balestro, G., Piana, F., Fioraso, G., Perrone, G. et Tallone, S. (2013) Sharing data and interpretations of geological maps via standardised metadata and geoportals. *Italian Journal of Geosciences*, vol. 132, n°2, p. 254-262.
- Bater, C. W. et Coops, N. C. (2009) Evaluating error associated with lidar-derived DEM interpolation. *Computers & Geosciences*, vol. 35, n°2, p. 289-300.
- Bonham-Carter, G. F. (1994) Introduction to GIS. *Geographic Information Systems for Geoscientists*, Elsevier, p. 1-23.
- Bouladon, J. (1964) Les gisements métallifères. *Géologie et principes de recherche* par P. Routhier, 1963. *Bulletin de Minéralogie*, vol. 87, n°1, p. 119-120.
- Breiman, L. (1996) Bagging predictors. *Machine Learning*, vol. 24, n°2, p. 123-140.
- Breiman, L. (2001) Random Forests. *Machine Learning*, vol. 45, n°1, p. 5-32.
- Carranza, E. J. M., Van Ruitenbeek, F. J. A., Hecker, C., Van der Meijde, M. et Van der Meer, F. D. (2008) Knowledge-guided data-driven evidential belief modeling of mineral prospectivity in Cabo de Gata, SE Spain. *International Journal of Applied Earth Observation and Geoinformation*, vol. 10, n°3, p. 374-387.
- Carranza, E. J. M. (2011) Geocomputation of mineral exploration targets. *Computers & Geosciences*, vol. 37, n°12, p. 1907-1916.
- Carranza, E. J. M. et Laborte, A. G. (2015) Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers and Geosciences*, vol. 74, p. 60-70.

- Carranza, E. J. M. et Laborte, A. G. (2016) Data-Driven Predictive Modeling of Mineral Prospectivity Using Random Forests: A Case Study in Catanduanes Island (Philippines). *Natural Resources Research*, vol. 25, n°1, p. 35-50.
- Chen, Y., Lu, L. et Li, X. (2014) Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. *Journal of Geochemical Exploration*, vol. 140, n°May, p. 56-63.
- Chen, T.-H., Shang, W., Jiang, Z. M., Hassan, A. E., Nasser, M. et Flora, P. (2014a) Detecting performance anti-patterns for applications developed using object-relational mapping. *Proceedings of the 36th International Conference on Software Engineering - ICSE 2014*, New York : ACM Press. New York, USA, p. 1001-1012.
- Chen, Y. (2015) Mineral potential mapping with a restricted Boltzmann machine. *Ore Geology Reviews*, vol. 71, p. 749-760.
- Chen, Y. et Wu, W. (2017) Mapping mineral prospectivity using an extreme learning machine regression. *Ore Geology Reviews*, vol. 80, p. 200-213.
- Corgne, S., Magagi, R., Yergeau, M. et Sylla, D. (2010) An integrated approach to hydro-geological lineament mapping of a semi-arid region of West Africa using Radarsat-1 and GIS. *Remote Sensing of Environment*, vol. 114, n°9, p. 1863-1875.
- Cortes, C. et Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, vol. 20, n°3, p. 273-297.
- Coulibaly, L. (1996) Interprétation structurale des linéaments par traitement d'images satellitaires : cas des sous-provinces d'Abitibi et d'Opatika. *Département de géographie et télédétection (Vol. Mémoire de)*. 106 p.
- Cruys, H. (1965) Prospection pour columbotantalite dans la région d'Issia. *Campagne juillet 1963 - avril 1965*. Abidjan, 102 p.
- Deveaud, S. (2015) Caractérisation de la mise en place des champs de pegmatites à éléments rares de type LCT : exemples représentatifs de la chaîne Varisque. *Thèse de doctorat, Université d'Orléans*, 349 p.
- Evangelidis, K., Ntouros, K., Makridis, S. et Papatheodorou, C. (2014) Geospatial services in the Cloud. *Computers and Geosciences*, vol. 63, p. 116-122.
- Feybesse, Jean-Louis ; Milési, J. P. ; Ouédraogo, M. F. ; Prost, A. E. (1990) La ceinture Protérozoïque inférieure de Boromo-Goren (Burkina Faso): un exemple d'inférence

entre deux phases transcurrentes éburnéennes. C.R. Acad. Sci., vol. 310, p. 1353-1360.

Freund, Y. (2009) A more robust boosting algorithm. Computer Science and Engineering, UCSD, p. 8.

Gabr, S., Ghulam, A. et Kusky, T. (2010) Detecting areas of high-potential gold mineralization using ASTER data. Ore Geology Reviews, vol. 38, n°1-2, p. 59-69.

Gkatzoflias, D., Mellios, G. et Samaras, Z. (2013) Development of a *Web* GIS application for emissions inventory spatial allocation based on open source software tools. Computers and Geosciences, vol. 52, p. 21-33.

Granek, J. et Haber, E. (2015) Data mining for real mining: A robust algorithm for prospectivity mapping with uncertainties. SIAM International Conference on Data Mining 2015, SDM 2015, p. 145-153.

Hashim, M., Ahmad, S., Johari, M. A. M. et Pour, A. B. (2013) Automatic lineament extraction in a heavily vegetated region using Landsat Enhanced Thematic Mapper (ETM+) imagery. Advances in Space Research, vol. 51, n°5, p. 874-890.

Hariharan, S., Tirodkar, S., Porwal, A., Bhattacharya, A. et Joly, A. (2017) Random Forest-Based Prospectivity Modelling of Greenfield Terrains Using Sparse Deposit Data: An Example from the Tanami Region, Western Australia. Natural Resources Research, vol. 26, n°4, p. 489-507.

He, H. et Garcia, E. A. (2009) Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, vol. 21, n°9, p. 1263-1284

Hechenbichler, K. et Schliep, K. (2004) Weighted k-Nearest-Neighbor Techniques and Ordinal Classification. Projektpartner, Munich, 16 p.

Jacob, R. J. ; Bluck, B. J. ; Ward, J. D. (1999) Tertiary-age diamondiferous fluvial deposits of the lower orange river valley, southwestern africa. Economic Geology, vol. 94, p. 749-755.

Ji, X., Huang, Z., Fang, Y., Chen, B. et Sui, Z. (2007) Weight of Evidence Method Based on SIG and Its Application on Evaluation on Mineral Resources. Journal of Jilin University: Earth Science Edition, vol. 1, n°37, p. 54-58.

- Joly, A., Porwal, A. et McCuaig, T. C. (2012) Exploration targeting for orogenic gold deposits in the Granites-Tanami Orogen: Mineral system analysis, targeting model and prospectivity analysis. *Ore Geology Reviews*, vol. 48, p. 349-383.
- Joly, A., Porwal, A., McCuaig, T. C., Chudasama, B., Dentith, M. C. et Aitken, A. R. A. (2015) Mineral systems approach applied to GIS-based 2D-prospectivity modelling of geological regions: Insights from Western Australia. *Ore Geology Reviews*, vol. 71, p. 673-702.
- Kereszturi, G., Pullanagari, R. R., Mead, S., Schaefer, L. N., Procter, J., Schleiffarth, W. K. et Kennedy, B. (2018) Geological Mapping of Hydrothermal Alteration on Volcanoes from Multi-Sensor Platforms. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, p. 220-223.
- King, G. et Zeng, L. (2001) Logistic Regression in Rare Events Data. *Political Analysis*, vol. 9, n°02, p. 137-163.
- Koudou, U. et Kouamé, F., (2010) Contribution of data ETM + de Landsat, de l'analyse multicritère et d'un SIG à identifier des secteurs à potentiel aquifère en zone de socle du bassin. *Photo Interprétation European Journal of Applied Remote Sensing*, vol. 46, n°3/4.
- Koudou, A., Assoma, T. V., Adiaffi, B., M, Y. T. A., Kouame, K. F. et Lasm, T. (2014) Analyses Statistique Et Geostatistique De La Fracturation Extraite De L ' Imagerie Asar Envisat Du Sud- Est De La Côte D'Ivoire. *Larhyss Journal*, ISSN 1112-3680, vol. n°20, n°20, p. 147-166.
- Landis, J. R. et Koch, G. G. (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics*, vol. 33, n°1, p. 159.
- Lanucara, S., Oggioni, A., Modica, G. et Carrara, P. (2017) Interoperable sharing and visualization of geological data and instruments: A proof of concept. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10407 LNCS, p. 584-599.
- Li, N. (2010) Textural and rule-based lithological classification of remote sensing Data, and geological mapping in Southwestern Prieska Sub- basin, Transvaal Supergroup, South Africa. *In: Thesis presented in at Z. Adiri et al. / Advances in Space Research (Text.PhDThesis)*. Thèse de doctorat, Ludwig-Maximilians-Universität München.

p132 p.

- Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D. et Brovelli, A. (2012) Multivoxel Pattern Analysis for fMRI Data: A Review. *Computational and Mathematical Methods in Medicine*, vol. 2012, p. 1-14.
- MathWorks (1994) *Statistics and Machine Learning Toolbox in Support Matlab Documentation*, The MathWorks Inc., Suisse,
<https://www.mathworks.com/help/stats/svmhyperplane.png>
- Mccuaig, T. C. et Sherlock R.L. (2017) Exploration Targeting. *Proceedings of Exploration*, vol. 17, p. 75-82.
- Melcher, F., Graupner, T., Gäbler, H.-E., Sitnikova, M., Henjes-Kunst, F., Oberthür, T., Gerdes, A. et Dewaele, S. (2015) Tantalum-(niobium-tin) mineralisation in African pegmatites and rare metal granites: Constraints from Ta-Nb oxide mineralogy, geochemistry and U-Pb geochronology The late-orogenic TNO mineralisation associated with A-type granites in the Eastern Desert of Egypt shares geochemical features with the anorogenic Younger Granites of Nigeria. *Ore Geology Reviews*, vol. 64, p. 667-719.
- Meloux, J. (1959) Étude des régions de Zuénoula, d'Issia et du Mont Goma. Rap. BRGM. Dakar, 32 p.
- Miraki, S., Zanganeh, S. H., Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H. et Pham, B. T. (2019) Mapping Groundwater Potential Using a Novel Hybrid Intelligence Approach. *Water Resources Management*, vol. 33, n°1, p. 281-302.
- Nguyen, T., Jones, S., Soto-Berelov, M., Haywood, A. et Hislop, S. (2018) A Comparison of Imputation Approaches for Estimating Forest Biomass Using Landsat Time-Series and Inventory Data. *Remote Sensing*, vol. 10, n°11, p. 1825.
- Nkono, C., Féménias, O., Lesne, A., Mercier, J.-C. et Demaiffe, D. (2013) Fractal Analysis of Lineaments in Equatorial Africa: Insights on Lithospheric Structure. *Open Journal of Geology*, vol. 03, n°03, p. 157-168.
- Noack, S., Knobloch, A., Etzold, S. H., Barth, A. et Kallmeier, E. (2014) Spatial predictive mapping using artificial neural networks. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-2, p. 79-86.

- Ortiz, J. M. et Emery, X. (2006) Geostatistical estimation of mineral resources with soft geological boundaries: {A} comparative study. *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 106, n°8, p. 577-584.
- Otukey, J. R. et Blaschke, T. (2010) Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, vol. 12, p. 527-531.
- Paganelli, F., Grunsky, E. C., Richards, J. P. et Pryde, R. (2003) Use of RADARSAT-1 principal component imagery for structural mapping: a case study in the Buffalo Head Hills area, northern central Alberta, Canada. *Canadian Journal of Remote Sensing*, vol. 29, n°1, p. 111-140.
- Papon, A., Lemarchand, R. (1971) Géologie et minéralisation du sud-ouest de la Côte d'Ivoire. Synthèse des travaux de l'opération SASCA (1962 - 1968). SODEMI, Mém. BRGM, Bull. Dir. Mines et Géol., vol. 80, 286 p.
- Peng, L., Niu, R., Huang, B., Wu, X., Zhao, Y. et Ye, R. (2014) Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology*, vol. 204, p. 287-301.
- Peters, J., Baets, B. De, Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. De et Huybrechts, W. (2007) Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling*, vol. 207, n°2-4, p. 304-318.
- Porwal, A., Carranza, E. J. M. et Hale, M. (2004) A Hybrid Neuro-Fuzzy Model for Mineral Potential Mapping. *Mathematical Geology*, vol. 36, n°7, p. 803-826.
- Porwal, A., González-Álvarez, I., Markwitz, V., Mccuaig, T. C. et Mamuse, A. (2010) Weights-of-evidence and logistic regression modeling of magmatic nickel sulfide prospectivity in the Yilgarn Craton, Western Australia. *Ore Geology Reviews, Mineral Prospectivity Analysis and Quantitative Resource Estimation*, vol. 38, n°3, p. 184-196.
- Ravikumar, G., Khaparde, S. A. et Pradeep, Y. (2013) CIM oriented database for topology processing and integration of power system applications. 2013 IEEE Power & Energy Society General Meeting, p. 1-5.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. et Chica-Rivas, M. (2015)

- Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, vol. 71, p. 804-818.
- Rombach, M. (1960) Rapport de synthèse de la minéralisation alluvionnaire de la région de Saïoua (Issia). Abidjan, 3 p.
- Saadi, N. M., Zaher, M. A., El-Baz, F. et Watanabe, K. (2011) Integrated remote sensing data utilization for investigating structural and tectonic history of the Ghadames Basin, Libya. *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, n°5, p. 778-791.
- Sable, K. A. et Wohl, E. (2006) The Relationship of Lithology and Watershed Characteristics to Fine Sediment Deposition in Streams of the Oregon Coast Range. *Environmental Management*, vol. 37, n°5, p. 659-670.
- Sadeghi, M., Billay, A. et Carranza, E. J. M. (2015) Analysis and mapping of soil geochemical anomalies: Implications for bedrock mapping and gold exploration in Giyani area, South Africa. *Journal of Geochemical Exploration*, vol. 154, p. 180-193.
- Sain, S. R. et Vapnik, V. N. (2006) The Nature of Statistical Learning Theory. *Technometrics*, vol. 38, n°4, p. 409.
- Scott, M. et Jones, M. (2014) Gestion des données géoscientifiques publiques Centre international de l'industrie minière pour le développement Industrie minière pour le développement: guide des pratiques australiennes., 20 p.
- Shabankareh, M. et Hezarkhani, A. (2017) Application of support vector machines for copper potential mapping in Kerman region, Iran. *Journal of African Earth Sciences, GIS-based mineral potential targeting*, vol. 128, p. 116-126.
- Silverman, B. W. (1986) Density estimation for statistics and data analysis. Chapman and Hall. 22 p.
- Silverman, B. W. et Jones, M. C. (2006) E. Fix et J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). *International Statistical Review / Revue Internationale de Statistique*, vol. 57, n°3, p. 233.
- Song, M., Xie, Z., Luo, X., Wu, L. et Li, C. (2010) Research of catalogue service system based on China geological survey information grid platform. 2010 18th International

- Conference on Geoinformatics, IEEE Geoinformatics 2010, p. 1-5.
- Sun, T., Chen, F., Zhong, L., Liu, W. et Wang, Y. (2019) GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geology Reviews*, vol. 109, p. 26-49.
- Tehrany, M. S., Pradhan, B., Mansor, S. et Ahmad, N. (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, vol. 125, p. 91-101.
- Tien Bui, D., Tuan, T. A., Klempe, H., Pradhan, B. et Revhaug, I. (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, vol. 13, n°2, p. 361-378.
- Toth, R., Ribault, J., Gentile, J., Sperling, D. et Madabhushi, A. (2013) Simultaneous segmentation of prostatic zones using Active Appearance Models with multiple coupled levelsets. *Computer Vision and Image Understanding*, vol. 117, n°9, p. 1051-1060.
- Touzet, C., Yuan, Y., Bian, F., Chen, Y., Wu, W., Xiong, Y., Zuo, R., Ji, X., Huang, Z., Fang, Y., Chen, B., Sui, Z., Noack, S., Knobloch, A., Etzold, S. H., Barth, A. et Kallmeier, E. (2017) An open sharing and interoperating platform for spatial information based on GIS *Web* services. *Ore Geology Reviews*, vol. 82, p. 5938-5941.
- Triboulet, C. et Feybesse, J.-L. (1998) Les metabasites birimiennes et archéennes de la région de Toulepleu-Ity (Côte-d'Ivoire): des roches portées à 8 kbar (≈ 24 km) et 14 kbar (≈ 42 km) au Paléoproterozoïque. *Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science*, vol. 327, n°1, p. 61-66.
- Vidal, M., Prost, A.E., Aric, G. et Lemoine, S. (1992) Présence d'un socle antérieur à une suture océanique du Birimien inférieur en Côte d'Ivoire (Afrique de l'Ouest). *C.R. Acad. Sci.*, vol. 2, n°315, p. 1085-1090.
- Vidal, M., Gumiaux, C., Cagnard, F., Pouclet, A. et Pichon, M. (2009) Evolution of a Paleoproterozoic "orogeny in the West African Craton" (Ivory Coast). *Tectonophysics*, vol. 477, n°4, p. 145-149.
- Walsh, S. J. et Mynar, F. (1986) Landsat digital enhancements for lineament detection. *Environmental Geology and Water Sciences*, vol. 8, n°3, p. 123-128.

- Weihua, D. (2009) Generating on-demand *Web* mapping through progressive generalization. 2008 International Workshop on Education Technology and Training and 2008 International Workshop on Geoscience and Remote Sensing, ETT and GRS 2008, IEEE, vol. 2, p. 163-166.
- Weissman, C. D. et Bobrowski, S. (2009) The design of the force.com multitenant internet application development platform, p. 889-896.
- Wu, L., Xue, L., Li, C., Lv, X., Chen, Z., Guo, M. et Xie, Z. (2015) A geospatial information grid framework for geological survey. PLoS ONE, vol. 10, n°12, p. 29.
- Xiong, Y. et Zuo, R. (2018) GIS-based rare events logistic regression for mineral prospectivity mapping. Computers & Geosciences, vol. 111, p. 18-25.
- Yao, A. B., Goula, B. T. A., Kane, A., Mangoua, O. M. J. et Kouassi, K. A. (2016) Cartographie du potentiel en eau souterraine du bassin versant de la Lobo (Centre-Ouest, Côte d'Ivoire): approche par analyse multicritère. Hydrological Sciences Journal, vol. 61, n°5, p. 856-867.
- Yousefi, M. et Carranza, E. J. M. (2015) Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping. Computers and Geosciences, vol. 74, p. 97-109.
- Yu, L., Porwal, A., Holden, E.-J. et Dentith, M. C. (2012) Towards automatic lithological classification from remote sensing data using support vector machines. Computers & Geosciences, vol. 45, n°Supplement C, p. 229-239.
- Zhang, X., Pazner, M. et Duke, N. (2007) Lithologic and mineral information extraction for gold exploration using ASTER data in the south Chocolate Mountains (California). ISPRS Journal of Photogrammetry and Remote Sensing, vol. 62, n°4, p. 271-282.
- Zhang, J. et Zhu, Y. (2015) A method based on graphic entity for visualizing complex map symbols on the *Web*. Cartography and Geographic Information Science, vol. 42, n°1, p. 44-53.
- Zhao, J., Chen, S. et Zuo, R. (2016) Identifying geochemical anomalies associated with Au-Cu mineralization using multifractal and artificial neural network models in the Ningqiang district, Shaanxi, China. Journal of Geochemical Exploration, vol. 16, p. 54-64.
- Zuluaga, M. C., Norini, G., Lima, A., Albanese, S., David, C. P. et De Vivo, B. (2017)

Stream sediment geochemical mapping of the Mount Pinatubo-Dizon Mine area, the Philippines: Implications for mineral exploration and environmental risk. *Journal of Geochemical Exploration*, vol. 175, p. 18-35.

Zuo, R. et Carranza, E. J. M. (2011) Support vector machine: A tool for mapping mineral prospectivity. *Computers and Geosciences, Geocomputation of Mineral Exploration Targets*, vol. 37, n°12, p. 1967-1975.